



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN LINGÜÍSTICA**

**El argumento de validación de la prueba escrita del examen de español
como lengua extranjera para el ámbito académico de la UNAM**

TESIS

que para optar por el grado de
DOCTOR EN LINGÜÍSTICA

PRESENTA
ARTURO MENDOZA RAMOS

TUTORA
DRA. ALINA MARÍA SIGNORET DORCASBERRO
Centro de Enseñanza de Lenguas Extranjeras

MIEMBROS DEL COMITÉ TUTOR
DRA. ROSA ESTHER DELGADILLO MACÍAS
DRA. MARTHA JURADO SALINAS
Centro de Enseñanza para Extranjeros

Ciudad Universitaria, Cd.Mx. marzo de 2017



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Con gran cariño y afecto, para José Luis Domínguez Carvajal

AGRADECIMIENTOS

A mi tutora, la Dra. Alina María Signoret Dorcasberro, por su confianza y apoyo incondicional a mi formación humana, académica y profesional. Agradezco de todo corazón el impulso y fortaleza que me brindó con afabilidad para poder llevar a cabo esta investigación.

A la Dra. Martha Jurado Salinas, por la dedicación, compromiso y entrega para que esta investigación saliera a la luz. Gracias por confiar siempre en mí y en mi trabajo.

A la Dra. Rosa Esther Delgadillo Macías, por el ánimo y el ahínco con el cual ha seguido mi formación académica en la lingüística desde sus estadios iniciales.

A la Dra. María del Carmen Contijoch Escontría y a la Dra. Laura García Landa, por sus pertinentes y minuciosas observaciones, las cuales enriquecieron significativamente el presente trabajo.

A Vincent Folny, por sus invaluable comentarios y sugerencias para conducir el cauce de mi investigación.

Al Posgrado en Lingüística, al CEPE, al Posgrado de la UNAM, alumnos no hispanohablantes, docentes del posgrado, correctores de pruebas escritas, codificadores y prestadores de servicio social: ¡Gracias por su tiempo, su participación y sus valiosas contribuciones!

A mis queridos padres, quienes me brindaron lo más valioso y elemental que un ser humano puede recibir: educación y valores.

A mis queridos hermanos y hermanas, quienes son ejemplo de vida en mi andar.

A José Luis Domínguez Carvajal y su querida familia, por todo el apoyo brindado durante estos años de trabajo y el cariño con el cual me acogieron.

A mi querida abuela, quien es muestra de entereza, tenacidad y perseverancia.

A mi familia y amigos, quienes enriquecen mi vida día a día.

"El trabajo de un intelectual no es moldear la voluntad política de otros; es, a través del análisis que hace en su propio campo, reexaminar evidencias y suposiciones, para sacudir las formas habituales de trabajo y pensamiento, para disipar las familiaridades convencionales, para reevaluar reglas e instituciones y para participar en la formación de una voluntad política (donde tiene su papel de ciudadano para actuar)".

— **Michel Foucault**

SINOPSIS

Los exámenes de certificación de lenguas extranjeras son cada vez más importantes como documentos probatorios del dominio de un idioma. En el contexto académico, estos exámenes son empleados para garantizar que los prospectos no nativohablantes serán capaces de acoplarse –en términos lingüísticos– a las exigencias académicas requeridas por las instituciones de educación superior. Sin embargo, diseñar un examen también exige que se conduzca un proceso de validación en aras de garantizar una evaluación y toma de decisiones justa y equitativa para los candidatos. Este tipo de exámenes conlleva implicaciones en la vida de los examinados, razón por la cual es menester recolectar evidencia que permita conocer su funcionamiento y garantizar que el examen refleja el uso de la lengua en un ámbito específico.

En el contexto hispanohablante, no existe aún ningún examen de lenguas con fines académicos que sea administrado internacionalmente. Por esta razón, las universidades hispanas encaran dificultades para valorar la competencia comunicativa de sus prospectos universitarios. De hecho, es recurrente que éstas confeccionen su propio examen y lo administren a sus estudiantes una vez que hayan comenzado sus estudios universitarios o a prospectos que residen permanentemente en el lugar donde cursarán sus estudios. Tal es el caso del Centro de Enseñanza para Extranjeros que, tras varios años de labor, introdujo en 2015 un nuevo examen para evaluar la competencia comunicativa de sus prospectos universitarios no hispanohablantes. Paralelamente, el advenimiento de este examen trajo consigo la necesidad conducir estudios para validarlo.

Dado que el proceso de validación de un examen es intrincado y laborioso, se decidió abordar exclusivamente una parte de éste: la prueba escrita. Sin soslayar la importancia de las demás competencias comunicativas que requieren de validación, se optó por validar la prueba escrita por dos razones. En primer lugar, el examen que se administraba anteriormente no contaba con dicha prueba y, en segundo lugar, la escritura representa una habilidad fundamental dentro de las competencias académicas con que debe contar un estudiante universitario de grado o de posgrado.

Abordando la noción de validez, desde la lógica argumentativa del modelo de Toulmin (1958) y desde el enfoque del argumento de validación de exámenes de Kane (2006, 2013), se diseñó el argumento de validación de la prueba escrita del examen de certificación de español con fines académicos que desarrolló el Centro de Enseñanza para Extranjeros. Esta noción de validación de exámenes ha sido recientemente adoptada en las investigaciones que se llevan cabo dentro del campo de la evaluación y certificación de exámenes de lenguas. El enfoque plantea que para aseverar que un examen –en este caso la prueba escrita del examen– es válido, se debe establecer un andamiaje de inferencias que permitan, mediante la evidencia, respaldar la validez de éste.

ÍNDICE

	página
Sinopsis.....	I
Índice.....	III
Lista de cuadros.....	IX
Lista de figuras.....	X
Lista de gráficas.....	XI
Lista de tablas.....	XII
Lista de siglas y acrónimos.....	XIII
Nomenclatura empleada en MFRM.....	XIV
Glosario de términos en MFRM.....	XV

CAPÍTULO 1

INTRODUCCIÓN

	1
1.1. La población universitaria no hispanohablante en el posgrado.....	3
1.2. La problemática.....	5
1.3. Antecedentes de la investigación.....	6
1.4. El EXELEAA.....	7
1.4.1. La PE del EXELEAA.....	7
1.5. El constructo de la PE del EXELEAA.....	8
1.5.1. El sustento teórico de la PE del EXELEAA.....	9
1.5.2. La evaluación de la PE del EXELEAA.....	11
1.6 Justificación.....	12
1.7 Objetivos de la investigación.....	13
1.8 Aportaciones de la investigación.....	14
1.9 Organización de la tesis.....	15

CAPÍTULO 2

LA VALIDEZ EN LOS EXÁMENES DE LENGUAS EXTRANJERAS

	17
2.1. Introducción.....	17
2.2. Los orígenes de la validez en los exámenes.....	18
2.3. La validez a finales del siglo XX y comienzos del siglo XXI.....	20
2.4. El modelo de argumentación de Toulmin (1958) y Toulmin, Rieke y Janik (1984).....	21
2.5. Los enfoques de argumentación aplicados a la evaluación de exámenes....	25
2.5.1. El enfoque basado en el argumento de validación de Bachman.....	25
2.5.2. El enfoque basado en el argumento de validación de Kane.....	28
2.5.2.1. El argumento de interpretación y de uso.....	31
2.5.2.1.1. Inferencia de la evaluación.....	31
2.5.2.1.2. Inferencia de generalización.....	31
2.5.2.1.3. Inferencia de extrapolación.....	32
2.5.2.1.4. Inferencia de toma de decisiones.....	32
2.5.3. El modelo de validación de Mislav.....	33
2.6. La disyuntiva entre evidencia y verdad.....	36
2.7. Discusión sobre los enfoques de argumentación para validar exámenes.....	37
2.8. Resumen.....	39

	página
CAPÍTULO 3	
EL ARGUMENTO DE VALIDACIÓN DE LA PE DEL EXELEAA	41
3.1. Introducción.....	41
3.2. Diseño del argumento de validación de la PE del EXELEAA.....	41
3.3. Diseño de la investigación.....	43
CAPÍTULO 4	
Estudio 1: La descripción del ámbito académico y las tareas de la PE del EXELEAA	49
4.1. Introducción.....	49
4.2. Marco teórico.....	49
4.2.1. La descripción del ámbito académico	50
4.2.1.1. Las habilidades de escritura académica	51
4.2.1.2. Los géneros discursivos a nivel posgrado	53
4.2.1.3. La importancia de la escritura académica a nivel posgrado	57
4.2.2. Las tareas de escritura de los ECLAA	60
4.2.2.1. Los tipos de tareas de escritura para los ECLAA	63
4.2.2.2. Las variables intrínsecas asociadas al tipo de tarea escrita	65
4.2.2.3. La selección de la tarea	65
4.2.2.3.1. <i>Las tareas discretas</i>	66
4.2.2.3.2. <i>Las tareas integradas</i>	70
4.2.2.4. Otras variables importantes	73
4.2.2.4.1. <i>El modo discursivo</i>	73
4.2.2.4.2. <i>La claridad en la consigna</i>	75
4.2.2.4.3. <i>La planeación de la tarea</i>	75
4.2.2.4.4. <i>El tiempo de escritura</i>	76
4.2.2.4.5. <i>El uso de diccionario</i>	77
4.2.2.4.6. <i>El modo de escritura</i>	78
4.3. Métodos e instrumentos.....	79
4.3.1. Los participantes	80
4.3.2. El instrumento	81
4.3.3. El procedimiento	81
4.3.4. El análisis	81
4.4. Resultados.....	82
4.4.1. El cuestionario sobre prácticas de evaluación docente	83
4.4.1.1. Resultados del área 1: Ciencias Físico-Matemáticas e Ingenierías	84
4.4.1.2. Resultados del área 2: Ciencias Biológicas, Químicas y de la Salud	85
4.4.1.3. Resultados del área 3: Ciencias Sociales	85
4.4.1.4. Resultados del área 4: Humanidades y Artes	86
4.4.2. Encuesta sobre la importancia de las habilidades de escritura académica	86
4.5. Discusiones, limitaciones y futuras investigaciones.....	89
4.5.1. La descripción del ámbito: la escritura dentro del contexto académico	89
4.5.2. La pertinencia de las tareas de la PE del EXELEAA	92
4.5.3. Limitaciones y futuras investigaciones	95
4.6. Conclusiones.....	95

Capítulo 5

página

Estudio 2: La evaluación y la generalización de los resultados de la PE del EXELEAA	97
5.1. Introducción.....	97
5.2. Marco teórico.....	98
5.2.1. Las rúbricas empleadas para evaluar las PE.....	98
5.2.1.1. El tipo de rúbricas	99
5.2.1.1.1. <i>Las rúbricas holísticas</i>	100
5.2.1.1.2. <i>Las rúbricas analíticas</i>	102
5.2.1.2. El proceso de construcción de una rúbrica	106
5.2.1.3. El proceso de validación de una rúbrica	108
5.2.2. Los evaluadores de PE	112
5.2.2.1. La consistencia en los resultados de los evaluadores	114
5.2.2.2. La capacitación de los evaluadores	118
5.2.2.2.1. <i>Las sesiones de capacitación</i>	120
5.2.2.3. El monitoreo de los evaluadores	123
5.2.2.4. El proceso de evaluación de textos	124
5.2.2.4.1. <i>El evaluador electrónico</i>	126
5.2.2.5. Variables que influyen en la consistencia de los evaluadores	129
5.2.2.5.1. <i>Los antecedentes del evaluador</i>	130
5.2.2.5.2. <i>La cognición del evaluador</i>	132
5.2.2.5.3. <i>Variables superfluas que influyen en la apreciación del texto</i>	135
5.3. Métodos e instrumentos.....	136
5.3.1 La calidad del proceso de evaluación	137
5.3.1.1. Primera parte: pilotaje con la primera versión de la rúbrica	138
5.3.1.1.1. <i>La muestra</i>	138
5.3.1.1.2. <i>Las tareas de la PE</i>	138
5.3.1.1.3. <i>La primera versión de la rúbrica analítica</i>	138
5.3.1.1.4. <i>Los evaluadores</i>	139
5.3.1.1.5. <i>El proceso de corrección</i>	139
5.3.1.1.6. <i>El análisis estadístico en MFRM</i>	140
5.3.1.1.6.1. <i>El programa estadístico FACETS</i>	140
5.3.1.1.6.2. <i>Antecedentes de MFRM</i>	143
5.3.1.1.6.3. <i>Descripción del modelo MFRM</i>	144
5.3.1.1.6.4. <i>Críticas al modelo MFRM</i>	145
5.3.1.2. Segunda parte: corrección con la segunda versión de la rúbrica	146
5.3.1.2.1. <i>La muestra</i>	146
5.3.1.2.2. <i>Las tareas de la PE</i>	146
5.3.1.2.3. <i>La segunda versión de la rúbrica analítica</i>	147
5.3.1.2.4. <i>Los evaluadores</i>	148
5.3.1.2.5. <i>El proceso</i>	148
5.3.1.2.6. <i>El análisis</i>	149
5.3.2. La percepción de los evaluadores sobre la rúbrica	149
5.3.2.1. Los participantes	149
5.3.2.2. El instrumento	149
5.3.3. El alcance de la rúbrica de la PE para evaluar habilidades de escritura académica	149
5.3.3.1. Los participantes	150
5.3.3.2. Los instrumentos	150
5.3.3.3. El proceso	150
5.4. Resultados.....	150
5.4.1. Las tareas de la prueba escrita	151
5.4.2. La calidad del proceso de evaluación mediante MFRM	156
5.4.2.1. Pilotaje: evaluación con la primera versión de la rúbrica	156
5.4.2.2. Modificación de la primera versión de la rúbrica analítica	166
5.4.2.3. Segunda parte: corrección con la segunda versión de la rúbrica	171
5.4.3. La percepción de los evaluadores sobre la rúbrica	189

	página
5.4.4. El alcance de la rúbrica de la PE para evaluar habilidades de escritura académica	182
5.5. Discusiones, limitaciones y futuras investigaciones	193
5.5.1. Inferencia de evaluación	193
5.5.1.1. La calidad de la rúbrica según los parámetros de MFRM	193
5.5.1.2. La percepción de los evaluadores sobre la rúbrica	194
5.5.1.3. El alcance de la PE para evaluar habilidades de escritura académica	196
5.5.2. Inferencia de generalización	197
5.5.3. Limitaciones y futuras investigaciones	198
5.6. Conclusiones	199

CAPÍTULO 6

Estudio 3: La extrapolación de los resultados de la PE del EXELEAA al “mundo real”	201
6.1. Introducción.....	201
6.2. Marco teórico.....	202
6.2.1. La correlación de un examen de lengua con las notas académicas	202
6.2.2. Las comparaciones textuales	205
6.2.3. La correlación del examen con criterios alternos: entrevistas y cuestionarios	209
6.3. Métodos e instrumentos.....	216
6.3.1. Los participantes	217
6.3.2. Primera parte: la correlación de la PE del EXELEAA con otros instrumentos de evaluación de escritura académica	219
6.3.2.1. La encuesta de autoevaluación de las habilidades de escritura académica	219
6.3.2.2. La entrevista semiestructurada	219
6.3.2.3. La rúbrica analítica para evaluar la entrevista	220
6.3.2.4. El análisis cuantitativo mediante el software SPSS	221
6.3.3. Segunda parte: la escritura académica de los estudiantes no hispanohablantes	221
6.3.3.1. La entrevista semiestructurada	221
6.3.3.2. El proceso de transcripción de las entrevistas	222
6.3.3.3. El análisis de los datos	223
6.3.3.4. El análisis del contenido	224
6.3.3.5. El proceso de categorización	225
6.3.3.6. El proceso de codificación	226
6.3.3.7. El proceso de interpretación	227
6.3.3.8. El software NVivo para el análisis cualitativo	227
6.3.3.9. Descripción del funcionamiento del programa NVivo	228
6.4. Resultados.....	237
6.4.1. Primera parte: la correlación de la PE del EXELEAA con otros instrumentos de evaluación de escritura académica	237
6.4.2. Segunda parte: la escritura académica de los estudiantes no hispanohablantes	242
6.4.2.1. Grupo intermedio bajo	243
6.4.2.2. Grupo intermedio alto	251
6.4.2.3. Grupo avanzado	264
6.4.2.4. Otros conceptos no incluidos en las categorías	278
6.5. Discusiones, limitaciones y futuras investigaciones.....	279
6.5.1. Primera parte: la correlación de la PE del EXELEAA con otros instrumentos de evaluación de escritura académica	280
6.5.2. Segunda parte: la escritura académica de los estudiantes no hispanohablantes	283

	página
6.5.2.1. El modelo de escritura académica en español como segunda lengua	283
6.5.2.1.1. <i>El proceso de escritura en español como lengua extranjera</i>	285
6.5.2.1.2. <i>La percepción de la escritura académica</i>	289
6.5.2.1.3. <i>Aspectos que favorecen la escritura académica en español</i>	290
6.5.2.1.4. <i>La reflexión sobre la mejora de la escritura académica</i>	291
6.5.3. Limitaciones y futuras investigaciones	291
6.6. Conclusiones.....	293

CAPÍTULO 7

DISCUSIONES Y CONCLUSIONES GENERALES	295
7.1. Discusiones generales.....	295
7.2. Conclusiones generales.....	302

REFERENCIAS	305
--------------------	-----

ANEXOS

Estudio 1	i
Anexo 1.1 Docentes de posgrado participantes en el estudio sobre las habilidades de escritura requeridas en el posgrado.....	i
Anexo 1.2 Cuestionario sobre la evaluación en el posgrado.....	iii
Anexo 1.3 Encuesta sobre las habilidades de escritura académica en el posgrado.....	iv
Anexo 1.4 Resultados del cuestionario sobre las prácticas de evaluación en el posgrado.....	viii
Anexo 1.5 Comentarios adicionales de los docentes de posgrado.....	xvii
Estudio 2	xxix
Anexo 2.1 Prueba escrita del EXELEAA (modelo 1).....	xxix
Anexo 2.2 Rúbrica para evaluar las 2 tareas que conforman la prueba de expresión escrita del EXELEAA (primera versión).....	xxx
Anexo 2.3 Matriz de evaluaciones con la primera versión de la rúbrica.....	xxxiii
Anexo 2.4 Estudiantes universitarios que participaron en la prueba escrita del EXELEAA.....	xxxiv
Anexo 2.5 Prueba escrita del EXELEAA (modelo 2).....	xxxvi
Anexo 2.6 Rúbrica para evaluar las 2 tareas que conforman la prueba de expresión escrita del EXELEAA (segunda versión).....	xxxvii
Anexo 2.7 Matriz de evaluaciones con la segunda versión de la rúbrica.....	xli
Anexo 2.8 Cuestionario sobre la rúbrica y el proceso para evaluar los textos del EXELEAA.....	xliii
Anexo 2.9 Resultados del cuestionario sobre la rúbrica y el proceso para evaluar los textos del EXELEAA.....	xlvi
Anexo 2.10 Habilidades de escritura que mide la segunda versión de la rúbrica.	li

Estudio 3	lii
Anexo 3.1 Estudiantes de posgrado de la UNAM que participaron en las entrevistas sobre escritura académica.....	lii
Anexo 3.2 Encuesta sobre el dominio de las habilidades de escritura académica en el posgrado.....	liii
Anexo 3.3 Guion de entrevista semiestructurada para estudiantes de posgrado no hispanohablantes de la UNAM.....	liv
Anexo 3.4 Rubrica para evaluar las entrevistas sobre la habilidad de escritura académica de los estudiantes no hispanohablantes del posgrado de la UNAM....	lvi
Anexo 3.5 Niveles de logro de la prueba escrita del Examen de Español para el Ámbito Académico.....	lvii

LISTA DE CUADROS

	Página
Cuadro 1.1: Esquema general del EXELEAA.....	7
Cuadro 2.1: Facetas de validez como una matriz progresiva.....	20
Cuadro 3.1: Esquema general de la investigación.....	44
Cuadro 4.1: Habilidades de escritura más importantes a nivel maestría.....	53
Cuadro 4.2: Tareas de escritura requeridas en el contexto universitario anglófono.....	54
Cuadro 4.3: Posgrados participantes.....	80
Cuadro 4.4: Jerarquización de las habilidades de escritura por parte de los docentes del posgrado de la UNAM.....	87
Cuadro 4.5: Comparativo entre las habilidades de escritura más importantes para los docentes de posgrado de la UNAM y de EEUU.....	89
Cuadro 5.1: Ventajas y desventajas de las rúbricas holísticas y analíticas.....	104
Cuadro 5.2: Efectos de los evaluadores.....	131
Cuadro 5.3: Datos de los examinados.....	147
Cuadro 5.4: Interpretación de los índices de ajuste.....	161
Cuadro 5.5: Competencia estratégica (primera versión de la rúbrica).....	167
Cuadro 5.6: Complejidad sintáctica (primera versión de la rúbrica).....	168
Cuadro 5.7: Puntuación (primera versión de la rúbrica).....	168
Cuadro 5.8: Registro académico (primera versión de la rúbrica).....	169
Cuadro 5.9: Registro académico (segunda versión de la rúbrica).....	169
Cuadro 5.10: Precisión gramatical (primera versión de la rúbrica).....	170
Cuadro 5.11: Precisión gramatical (segunda versión de la rúbrica).....	170
Cuadro 5.12: La evaluadora 4 es sumamente indulgente.....	172
Cuadro 5.13: La evaluadora 4 es sumamente severa.....	173
Cuadro 5.14: Examinado con mayor índice de desajuste.....	180
Cuadro 5.15: Examinado con menor índice de desajuste.....	181
Cuadro 5.16: Examinado con altos puntajes en ortografía.....	184
Cuadro 5.17: Resultado del juicio de expertos.....	192
Cuadro 6.1: Distribución de estudiantes en grupos de acuerdo con el puntaje obtenido en la PE del EXELEAA.....	238
Cuadro 6.2: Distribución de estudiantes en grupos de acuerdo con el puntaje medio obtenido en la entrevista.....	239
Cuadro 6.3: Resultados de los puntajes medios obtenidos en la PE del EXELEAA, la encuesta y la entrevista.....	240
Cuadro 6.4: Estudiantes clasificados en grupos de acuerdo al puntaje que obtuvieron en la entrevista.....	242

LISTA DE FIGURAS

	Página
Figura 1.1: Componentes del marco teórico organizados en la rúbrica.....	12
Figura 2.1: Modelo de argumentación de Toulmin.....	22
Figura 2.2: Argumento de uso en la evaluación (AUA).....	27
Figura 2.3: Procedimiento de medición y argumento para la interpretación de rasgos.....	30
Figura 2.4: Modelo basado en el razonamiento para la evaluación de lenguas....	35
Figura 3.1: Argumento de validación de la PE del EXELEAA.....	42
Figura 5.1: Marco conceptual y psicométrico de los factores relevantes en la evaluación del desempeño mediada por los evaluadores.....	116
Figura 5.2: Ejemplo de texto donde la caligrafía incide negativamente de forma visual.....	153
Figura 5.3: Los textos en computadora se perciben de longitud menor en comparación con los escritos a mano.....	154
Figura 5.4: La mala caligrafía dificulta la lectura de los textos.....	155
Figura 5.5: Mapa de variabilidad del pilotaje: primera versión de la rúbrica.....	158
Figura 5.6: Mapa de variabilidad de corrección con la segunda versión de la rúbrica.....	176
Figura 6.1: Organización de las transcripciones en carpetas.....	229
Figura 6.2: Categorías o nodos identificados para las entrevistas.....	231
Figura 6.3: Selección de la información dentro de una categoría o nodo.....	232
Figura 6.4: Ejemplo de categoría con altas correlaciones entre codificadores.....	233
Figura 6.5: Ejemplo de categoría con bajas correlaciones entre codificadores....	234
Figura 6.6: Ejemplo de alta correlación (1) entre codificadores.....	235
Figura 6.7: Ejemplo de correlación 0 entre codificadores.....	236
Figura 6.8: Ejemplo de correlación negativa entre codificadores.....	236
Figura 6.9: Modelo de escritura académica en español como segunda lengua...	284

LISTA DE GRÁFICAS

	Página
Gráfica 1.1: Estudiantes no hispanohablantes inscritos en el posgrado de la UNAM en el período 2014-1 (maestría).....	4
Gráfica 1.2: Estudiantes no hispanohablantes inscritos en el posgrado de la UNAM en el período 2014-1 (doctorado).....	4
Gráfica 1.3: Programas de posgrado con mayor matrícula de estudiantes no hispanohablantes en 2014 (maestría).....	4
Gráfica 1.4: Programas de posgrado con mayor matrícula de estudiantes no hispanohablantes en 2014 (doctorado).....	4
Gráfica 4.1: Clasificación de familias de géneros textuales.....	56
Gráfica 4.2: Prácticas de evaluación docente en el Posgrado de la UNAM.....	83
Gráfica 5.1: Medias de los evaluadores para el examinado 47.....	172
Gráfica 5.2: Medias de los evaluadores para el examinado 8.....	173
Gráfica 5.3: Sesgos e interacciones entre evaluadores y el examinado 40.....	187
Gráfica 5.4: Sesgos e interacciones entre evaluadores y el examinado 12.....	187
Gráfica 5.5: Sesgos e interacciones entre evaluadores y examinados.....	188

LISTA DE TABLAS

	Página
Tabla 5.1: Resumen de medición de datos.....	157
Tabla 5.2a: Reporte de medición de los estudiantes con la primera versión de la rúbrica.....	164
Tabla 5.2b: Reporte de medición de los evaluadores con la primera versión de la rúbrica.....	165
Tabla 5.2c: Reporte de medición de las tareas con la primera versión de la rúbrica.....	165
Tabla 5.2d: Reporte de medición de las categorías de la primera versión de la rúbrica.....	166
Tabla 5.3: La evaluadora 4 es asistemática (segunda versión de la rúbrica).....	171
Tabla 5.4a: Reporte de medición de los candidatos con la segunda versión de la rúbrica.....	178
Tabla 5.4b: Reporte de medición de los evaluadores con la segunda versión de la rúbrica.....	182
Tabla 5.4c: Reporte de medición de las tareas con la segunda versión de la rúbrica.....	183
Tabla 5.4d: Reporte de medición de las categorías de la segunda versión de la rúbrica.....	184
Tabla 5.5: Reporte de sesgos entre evaluador y examinado con la segunda versión de la rúbrica.....	185
Tabla 6.1: Correlaciones de Pearson (examen, encuesta y entrevista).....	241

LISTA DE SIGLAS Y ACRÓNIMOS

AERA – *American Educational Research Association*
AES – *Automated Essay Scoring*
APA – *American Psychological Association*
AUA – *Assessment Use Argument*
CAEL – *Canadian Academic English Language*
CENEVAL – Centro Nacional de Evaluación
CEPE – Centro de Enseñanza para Extranjeros
DELE – Diploma de Español como Lengua Extranjera
ECLAA – Examen de Certificación de Lengua para el Ámbito Académico
ECLAE – Examen de Certificación de Lengua para Ámbitos Específicos
ECLAG – Examen de Certificación de Lengua para Ámbitos Generales
EXELEAA – Examen de Español como Lengua Extranjera para el Ámbito Académico
EXHALING – Examen de Habilidades Lingüísticas
EPLA – Examen de Posesión de la Lengua Española
GEPT – *General English Proficiency Test*
GMAT – *Graduate Management Admission Test*
GRE – *Graduate Record Examination*
IUA – *Interpretation Use Argument*
IELTS – *International English Language Testing System*
LSM – Lingüística Sistemática Funcional
MCER – Marco Común Europeo de Referencia
MELAB – *Michigan English Language Assessment Battery*
MFRM- *many-facet Rasch measurement*
NCME: *National Council on Measurement in Education*
PAF – *Principal Axis Factoring*
PCA – *Principal Component Analysis*
PE – Prueba Escrita
TestDaF – *Test für Deutsch als Fremdsprache*
TOEFL CBT – *Test of English as a Foreign Language, Computer-Based Test*
TOEFL iBT – *Test of English as a Foreign Language, Internet-Based Test*
TOEFL PBT – *Test of English as a Foreign Language, Paper-Based Test*
TOEIC – Test of English for International Communication
SAT – *Scholastic Aptitude Test*
UNAM – Universidad Nacional Autónoma de México

NOMENCLATURA EMPLEADA EN MFRM

Lógito – log-odd unit o logaritmo natural

Cat – categorías

Score – puntaje

Exp – puntaje esperado

Resd – residual

StRes – residual estandarizado

Measr – medición en lógitos

S.E. – error estándar:

Infit MnSq – media ponderada de los cuadrados de las diferencias estandarizadas

OutfitMnSq – media no ponderada de los cuadrados de las diferencias estandarizadas

Point-biserial – puntos biserials

PtMea – medida de puntos de correlación

PtExp – medida de puntos de correlación esperados por el modelo

Zstd – medias cuadradas estandarizadas

Root mean-square error – promedio estadístico del error estándar de las mediciones

Adj (True) SD – verdadera desviación estándar

Separation – estratos de medición distinguibles entre las mediciones

Exact obs – observaciones exactas

Agree exp – observaciones esperadas por el modelo

GLOSARIO DE TÉRMINOS EN MFRM¹

Categoría: Niveles de desempeño o manifestación considerados en un instrumento, como parte de un formato de respuesta o de una guía de observación.

Desviación estándar: Raíz cuadrada del promedio de las diferencias cuadráticas, obtenidas entre las medidas en lógitos y la media.

Error estándar: Cantidad estimada respecto de la precisión de una medida o calibración. Si se suma y resta el error estándar a una medida o calibración en lógitos se obtiene un intervalo, cuyos límites señalan la distancia requerida para dar significado a la diferencia respecto de otra medida o calibración.

Escala: Representación cuantitativa de una variable.

Infit: Estadístico de ajuste con información ponderada que se enfoca al comportamiento general de un ítem o de una persona. Se calcula con el promedio ponderado de las desviaciones (o diferencias) cuadráticas estandarizadas entre el desempeño observado y el esperado. El estadístico que se presenta en forma gráfica y tabular en el análisis de Rasch es la media cuadrática normalizada. Detecta desajustes en las desviaciones cerca de la zona de medición del ítem. El nombre "infit" es una abreviatura de "ajuste con información ponderada", por ello no se traduce al castellano. No debe traducirse como "ajuste interno". En algunos países latinoamericanos se denomina también como "ajuste próximo" que puede parecer más apropiado.

Lógito: Unidad de medida usada en el modelo de Rasch para la calibración de los ítems y la medida de las personas. Se trata de una transformación del logaritmo natural del momio de la probabilidad de una respuesta correcta.

Mapa: Diagrama de barras (o histograma combinado) que muestra en dos partes de un mismo gráfico, a lo largo de la variable, la frecuencia y dispersión de los agentes (generalmente del lado derecho) contra la frecuencia y dispersión de los objetos (generalmente del lado izquierdo).

Medida: Ubicación de la variable latente (generalmente en unidades de lógitos). En el modelo de Rasch la medida de las personas se denomina "habilidad" y la medida de los ítems se denomina "dificultad".

Outfit: Estadístico de ajuste sensible a los casos atípicos, que permite determinar eventos poco usuales que ocurren de modo inesperado. Se determina por el promedio de las desviaciones (o diferencias) cuadráticas estandarizadas entre el desempeño observado y el esperado. En las gráficas y tablas de Rasch se emplean las medias cuadráticas normalizadas no ponderadas, para que queden centradas simétricamente respecto de cero. Detecta desajustes en las desviaciones lejos de la zona de medición del ítem. El nombre "outfit" es una forma abreviada de "ajuste sensible a los casos atípicos (outliers)", por ello no se traduce al castellano. No puede interpretarse como "ajuste externo". En algunos países latinoamericanos se denomina también como "ajuste lejano" que puede considerarse más apropiado.

Puntaje: Valores numéricos asignados a las respuestas cuando se suman para producir la calificación de un agente u objeto.

Residuales: Diferencia entre los valores observados y los valores esperados.

Respuesta esperada: Respuesta establecida teóricamente por el modelo de Rasch en el análisis de un agente u objeto.

¹ Terminología tomada de Agustín Tristán disponible en el portal de Rasch:

<http://www.rasch.org/rmt/glosario.htm>

Última consulta 25 de enero de 2017.

Respuesta observada: Respuesta actual obtenida por un objeto ante un agente dado.

Variable: Es la idea o definición conceptual de lo que se desea medir. Una variable se define por medio de ítems o agentes de medida, usados para hacer explícitas sus manifestaciones o respuestas.

Variable latente: Variable que no puede observarse directamente y que se estima en función de manifestaciones observables, que se suponen representativas de dicha variable.

CAPÍTULO 1

INTRODUCCIÓN

“If validity is to be considered the most fundamental consideration in developing and evaluating tests, it needs to address the suitability of the test for its intended function”.

(Kane, 2013, p. 62)

La globalización y el rápido crecimiento de los medios de comunicación han favorecido la migración por diversos motivos, entre los que destacan los laborales, académicos y familiares. Una de las consecuencias más conspicuas de dicho fenómeno es la necesidad de evaluar –tanto de manera general como específica en distintos ámbitos– la competencia¹ comunicativa de quienes arriban a países con lenguas distintas a la materna. Generalmente, dicha evaluación se efectúa mediante exámenes de lengua extranjera diseñados para determinar el dominio de un sustentante; de ahí que, los exámenes de lengua jueguen un papel primordial en la vida de quienes desean insertarse dentro de una comunidad lingüística ajena a la propia.

En general, los exámenes que certifican el dominio de una lengua pueden ser de dos tipos: generales o específicos. Los primeros son aquellos en los cuales no existe un propósito definido para la evaluación, razón por la cual tienden a cubrir un espectro más amplio de habilidades lingüísticas y generalmente son diseñados a partir de bases teóricas. En cambio, los exámenes diseñados para ámbitos específicos responden a necesidades particulares que deben verse reflejadas en el instrumento de evaluación. Unos de los exámenes de lengua con fines específicos más socorridos son aquellos destinados al ámbito académico. Indiscutiblemente, el interés de las universidades por establecer convenios internacionales de intercambio académico ha estimulado copiosamente la movilidad estudiantil. Esta movilidad se percibe también en quienes desean cursar estudios de posgrado en países en donde el medio de enseñanza es distinto al de la lengua materna.

¹ Beneitone, P., Esquetini, C., González, J., Marty, M., Siufi, G. y Wagenaar, R. (2007, p. 35) definen competencias como “las capacidades que todo ser humano necesita para resolver, de manera eficaz y autónoma, las situaciones de la vida. Se fundamentan en un saber profundo, no sólo saber qué y saber cómo, sino saber ser persona en un mundo complejo, cambiante y competitivo”.

Por esta razón, resulta imprescindible que los estudiantes cuenten con certificados que avalen su nivel de dominio de una lengua extranjera. En el caso del español, a diferencia del inglés, por ejemplo, no existen exámenes de lengua –destinados al ámbito académico– administrados internacionalmente. Por esta razón, es común que las universidades, como es el caso de la Universidad Nacional Autónoma de México (UNAM), diseñen sus propios exámenes de lengua para evaluar la competencia lingüística de sus prospectos universitarios no hispanohablantes. Debido a que los exámenes de lengua con fines académicos conllevan importantes implicaciones en la vida de los examinados, se les conoce como exámenes de alto impacto². Esto quiere decir que aprobar o no un examen repercute en las decisiones de admisión de un programa académico. Por esta razón, es necesario garantizar la validez de éstos y también promover una evaluación justa y equitativa; puesto que una mala decisión podría excluir a un candidato a cursar estudios universitarios o de forma contraria, aceptar a un prospecto cuya competencia comunicativa no sea suficiente para que se acople a las necesidades lingüísticas requeridas. Esta validación no sólo genera confianza en las distintas partes interesadas (p.ej., estudiantes, docentes, directores y organizaciones educativas), sino que también son de gran utilidad para quienes se encuentran en la posición de tomar decisiones respecto del desempeño de los examinandos.

En el caso de los exámenes de certificación de lengua para el ámbito académico (de aquí en adelante, ECLAA), se consideran pruebas y tareas que miden determinadas habilidades de la lengua encaminadas a evaluar el uso de la lengua en situaciones que podrían ocurrir dentro del contexto universitario. Por ello, éstos generalmente evalúan las cuatro habilidades³ y, de manera directa o indirecta, la gramática. Este esquema es típico de los exámenes de lenguas, aunque también es cierto que recientemente se están construyendo exámenes que evalúan las

² Un examen de alto impacto es aquel cuyo resultado es empleado para tomar decisiones importantes que conllevan consecuencias positivas o negativas en quienes los sustentan. A diferencia de los exámenes a gran escala, los de alto impacto pueden o no ser administrados a grandes poblaciones. Por ejemplo, en el contexto educativo, los resultados de los exámenes de aptitud, certificación y dominio son frecuentemente empleados por las partes interesadas como criterios de egreso/ingreso tanto en la educación media superior como superior.

³ Bachman y Palmer (2010, p. 55) critican el término empleado para definir las “cuatro habilidades”, pues mencionan que la única razón que subyace a esta división se debe al canal de transmisión (auditivo y visual) así como al modo (productivo y receptivo). Sin embargo, los autores mencionan que esta nomenclatura es limitada, puesto que hay múltiples casos en los que la línea no es tan clara. Por ejemplo, mencionan los autores que en caso del chat, se trata en realidad de comunicación oral pero en modo escrito. Entonces, los autores optan por el término “uso de la lengua”, mismo que definen como “la creación o interpretación de un significado intencionado en un discurso por un individuo, o como la negociación dinámica e interactiva de mensajes intencionados entre dos o más individuos en una situación particular”.

habilidades de forma integrada; es decir, combinan la comprensión auditiva o de lectura con la expresión oral y/o escrita. De igual forma, también es común que se evalúe la gramática integrada con la producción oral y escrita.

Ahora bien, en los exámenes de lengua para fines específicos, como los ECLAA, el proceso de validación no se circunscribe exclusivamente a las cualidades internas de éstos, sino también a su validación externa. Esta validación externa hace referencia al conocimiento del ámbito para el cual estamos evaluando, por una parte, mientras que por otra, resulta necesario investigar si, efectivamente, la competencia comunicativa impacta de la forma prevista en el futuro desempeño del examinado en el uso de la lengua en el contexto para el cual se lo evalúa. En suma, el proceso de validación no solamente se restringe a las cualidades internas de un examen, sino también a aquellas externas.

1.1. LA POBLACIÓN UNIVERSITARIA NO HISPANOHABLANTE EN EL POSGRADO

La UNAM cuenta con 52 programas de maestría, 35 de doctorado y más de 100 especialidades, entre las que destacan 78 especialidades médicas y 12 odontológicas⁴. Estos posgrados se encuentran divididos en cuatro áreas del conocimiento; a saber: Ciencias Físico-Matemáticas e Ingenierías; Ciencias Biológicas, Químicas y de la Salud; Ciencias Sociales; y Humanidades y Artes. En 2014, el total de estudiantes de posgrado no hispanohablantes se encontraba distribuido en 48 alumnos de maestría y 56 de doctorado⁵. La Gráficas 1.1 y 1.2 muestran la distribución de estudiantes no hispanohablantes inscritos en 2014 en programas de maestría o doctorado en la UNAM. Como se puede apreciar en las Gráficas 1.1 y 1.2, aproximadamente el 50% de los estudiantes no hispanohablantes se encontraban en el área de Humanidades y Artes, mientras que alrededor de un 30% se ubicaba en las Ciencias Sociales. Las áreas de Ciencias Físicas, Matemáticas e Ingenierías, así como las Ciencias Biológicas y Químicas, representaban tan sólo un 10% de estudiantes en maestría, y cerca de un 20% en doctorado. Lo anterior da cuenta de la importancia de los posgrados en Humanidades y Artes para los estudiantes no hispanohablantes que desean ingresar a la UNAM.

⁴ Consúltese el informe del Posgrado de la UNAM en cifras de 2015.

Recuperado del portal:

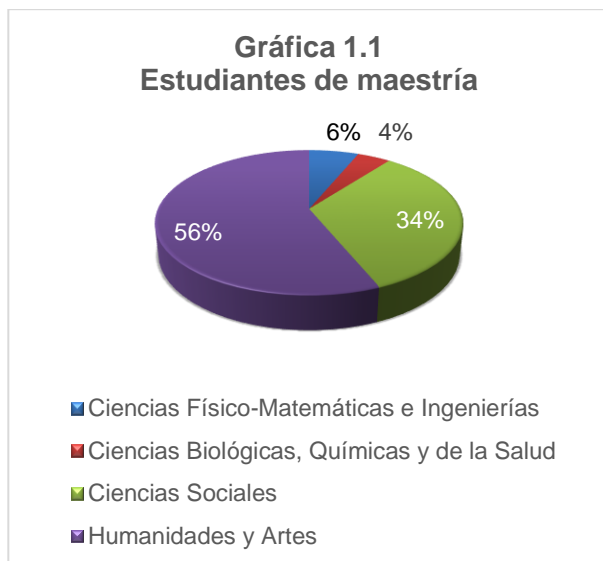
http://www.posgrado.unam.mx/publicaciones/otras/El_posgrado_en_cifras.pdf

Última consulta el 20 de diciembre de 2016.

⁵ Información proporcionada por el posgrado de la UNAM

Gráficas 1.1 y 1.2:

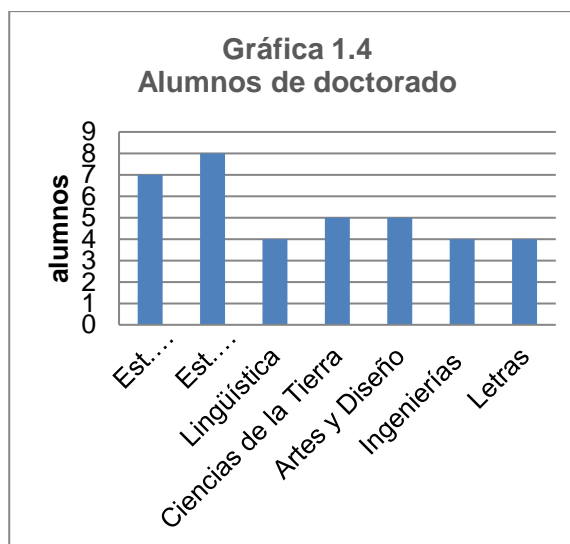
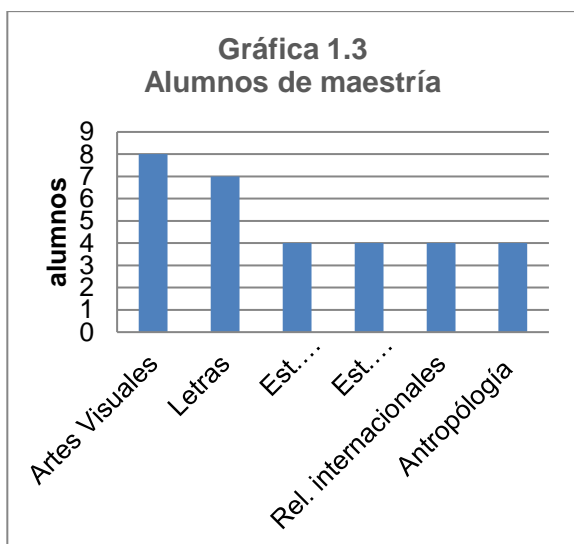
Estudiantes no hispanohablantes inscritos en el posgrado de la UNAM en el período 2014-1



En un análisis más fino, se puede apreciar que la distribución de estudiantes se encontraba concentrada en unos cuantos programas. Las Gráficas 1.3 y 1.4 muestran los programas de maestría y de doctorado con mayor matrícula de estudiantes no hispanohablantes. Estos programas, consistentes tanto en maestría como en doctorado, se ubican en las Ciencias Sociales y en las Humanidades y Artes, aunque destaca un posgrado perteneciente a las Ciencias Físico-Matemáticas e Ingenierías: Ciencias de la Tierra.

Gráficas 1.3 y 1.4:

Programas de posgrado con mayor matrícula de estudiantes no hispanohablantes en 2014



1.2. LA PROBLEMÁTICA

El examen de posesión de la lengua española (EPL), que hasta 2015 se utilizaba como criterio de ingreso o de egreso de los aspirantes no hispanohablantes a la UNAM, constaba de cuatro pruebas: comprensión auditiva, comprensión de lectura y vocabulario, estructuras y expresión oral, pero no de una prueba escrita (de aquí en adelante, PE). Sin embargo, el EPL no fue diseñado con fines meramente académicos, ni tampoco se llevaron cabo estudios de validación interna ni externa que pudieran respaldar su pertinencia y funcionamiento. Por esta razón, con la iniciativa del Centro de Enseñanza para Extranjeros (CEPE) y la colaboración de la Universidad de Guadalajara y la Universidad de Costa Rica, se decidió comenzar el diseño de un nuevo examen de español con fines académicos en 2010. Tras varios pilotajes que tuvieron lugar en 2013 y 2014 como parte de la presente investigación, se condujeron los análisis estadísticos necesarios que permitieron que el Examen de Español como Lengua Extranjera para el Ámbito Académico (de aquí en adelante, EXELEAA) remplazara al EPL.

Una vez que los exámenes son construidos es necesario darles un constante seguimiento que permita dar cuenta de su funcionamiento. A este seguimiento se le conoce como proceso de validación, y una de sus características fundamentales es su carácter iterativo; es decir, se trata de un proceso que nunca culmina. Fulcher y Davidson (2009) hacen una analogía entre un examen y un edificio. Los autores mencionan que un edificio se diseña y construye pensando en el uso que se le va a dar –igual que lo que sucede con un examen–. Asimismo, los autores agregan que una vez que se encuentra construido el edificio, se le debe dar un mantenimiento constante para que éste siempre se encuentre en óptimas condiciones de funcionalidad. Siguiendo esta analogía, los autores mencionan que los exámenes deben tener un seguimiento –denominado por los autores *retrofit*, que significa modernizar o actualizar– que permita eventualmente hacer cambios o actualizaciones de éstos, incluso sustituirlos por otros cuando sea necesario.

Por esta razón, dado que se trata de un nuevo examen, resulta ineludible realizar investigaciones que permitan garantizar su validez. En este caso, dado que la versión anterior no contaba con una PE de respuesta construida, se optó por validar esta prueba. Adicionalmente, el dominio de la escritura representa una de las habilidades primordiales dentro del contexto universitario.

1.3. ANTECEDENTES DE LA INVESTIGACIÓN

Actualmente, existen numerosos estudios que se han llevado a cabo para validar las distintas pruebas de las que constan los ECLAA. Destacan, sobre todo, los que continuamente se realizan en el contexto anglosajón con los exámenes del *Test of English as a Foreign Language* (TOEFL) y el *International English Language Testing System* (IELTS). Estos dos son los ECLAA que se administran a mayor escala a nivel mundial (con cerca de dos millones de examinados cada año). Es muy frecuente, sin embargo, que los estudios se dirijan a la validación interna de los exámenes, mas no a su validación externa.

En cuanto a la validación interna del examen, como se podrá ver en el Capítulo 5, Estudio 2 del presente trabajo, existen vastas investigaciones en relación con los examinados, las tareas, las escalas de evaluación empleadas y los evaluadores. En cuanto a las tareas, destacan aquellos estudios para ver cómo influyen diversas variables en el desempeño de los estudiantes; a saber: el uso del diccionario, el modo discursivo de la tarea, el modo de escritura, la planeación y la longitud de los textos, si la tarea es integrada o discreta, o la misma selección de la tarea. En cuanto a las escalas de corrección o rúbricas, también existen copiosos estudios que dan cuenta de la diferencia entre el tipo de escala: analítica u holística. En relación con los evaluadores, por ejemplo, hay trabajos que dan cuenta de las variables intrínsecas que determinan el tipo de enfoque del evaluador (p.ej., su antecedente académico, su estilo cognitivo, su conocimiento de otras lenguas extranjeras, entre otros). Asimismo, se han hecho trabajos que permitan evidenciar los índices de severidad o de indulgencia de los evaluadores. Éstos resultan relevantes para saber qué tan severo o indulgente es un evaluador al momento de evaluar.

En el caso de la validación externa del examen, como se podrá ver en el Capítulo 4, Estudio 1 y en Capítulo 6, Estudio 3 de la presente investigación, grosso modo se pueden dividir en estudios descriptivos del dominio para el cual se evalúa y estudios que evidencien el desempeño futuro de los examinados en el “mundo real⁶”. Los estudios descriptivos del ámbito para el cual deseamos evaluar permiten evidenciar la necesidad de construcción de un examen, pero también sirven para validar las tareas que se incluyen en él. Por otro lado, los trabajos que involucran el desempeño posterior de los examinados, conocidos previamente como validez predictiva, permiten dar cuenta si, efectivamente, el resultado obtenido por un

⁶ En la evaluación de lenguas, se emplea el término *mundo real* entre comillas para hacer la distinción entre la situación del uso de la lengua en un examen y en el ámbito para el cual se evalúa.

estudiante en un examen correlaciona positivamente con lo que éste es capaz de hacer en el ámbito para el cual fue evaluado.

Sin embargo, la gran mayoría de los estudios antes descritos se encuentran insertos en un contexto anglosajón en donde la cantidad de estudiantes que son evaluados anualmente permite que se conduzcan investigaciones que robustezcan la validez del examen. En otras lenguas, no obstante, los estudios son mucho más escasos. En el caso del español, debido a que aún no existe ningún ECLAA que sea administrado a gran escala, faltan estudios que se conduzcan en este sector.

1.4. EL EXELEAA

El EXELEAA consta de cinco secciones: competencia lingüística, comprensión de lectura, comprensión auditiva, expresión oral y expresión escrita. El Cuadro 1.1 muestra la duración total del examen (aproximadamente 3 horas 10 minutos) y la ponderación de cada una de las pruebas.

Cuadro 1.1: Esquema general del EXELEAA

Pruebas	Duración en minutos	No. de reactivos	Porcentaje
Competencia lingüística	15	20	20%
Comprensión de lectura	50	40	20%
Expresión oral	10		20%
Expresión escrita	70		20%
Comprensión auditiva	45	30	20%
Total	3 hrs. 10 min.	90	100%

1.4.1. LA PE DEL EXELEAA

La PE dura 70 minutos y contiene dos tareas. En el Anexo 2.1 se muestra un ejemplo de las tareas que tienen que resolver los candidatos que sustentan el EXELEAA. En la primera, se le solicita al sustentante que redacte un texto de por lo menos 150 palabras a partir de información no verbal. El objetivo de esta tarea es describir y contrastar información a partir de gráficos. En la segunda, se pide la redacción de un ensayo con una extensión de por lo menos 250 palabras. El objetivo de esta segunda tarea es sustentar un punto de vista. La tarea consiste en que el sustentante desarrolle un tema con base en una consigna planteada en temas controvertidos de la actualidad. El propósito de la PE del EXELEAA es evaluar las

habilidades de escritura académica que posee un candidato no hispanohablante que desea ingresar al ciclo universitario.

1.5. EL CONSTRUCTO DE LA PE DEL EXELEAA

Antes de describir el constructo de la PE EXELEAA, se explicará la noción de dicho concepto. El constructo se refiere a los antecedentes sobre los cuales se sustenta el diseño de un examen. En general, el constructo de un ECLAA parte de las competencias comunicativas estipuladas en los planes curriculares a nivel universitario. Asimismo, cuando se desea evaluar la competencia comunicativa en lengua extranjera, se emplean marcos de referencia que hacen alusión al desempeño de los candidatos con respecto a las habilidades productivas (expresión oral y escrita) y receptivas (comprensión de lectura y comprensión auditiva) de la lengua.

Bachman (2006) menciona que el constructo hace referencia a la entidad de interés; es decir, que “cuando un investigador observa algún fenómeno del mundo real, generalmente lo hace porque desea describir, inducir o explicar algo con base en la observación” (p. 183). En este sentido, el constructo en la evaluación de lenguas es útil, pues permite definir el uso de la lengua de acuerdo con la situación comunicativa y con el contexto en el cual se encuentra inserto el uso de la lengua. El autor agrega que actualmente, en el caso de la evaluación de lenguas, generalmente se concibe el constructo desde una perspectiva interaccionista en la cual se percibe una relación dialógica entre el usuario de la lengua y el contexto en el cual se la emplea. En cuanto al origen del constructo, el autor afirma que éste proviene del punto a partir del cual el investigador desea observar. La pregunta clave que el investigador debe plantearse con respecto al constructo, continúa Bachman, se refiere al grado de especificidad deseado. Bachman y Palmer (2010) definen el constructo de la siguiente manera:

For our purposes, we can consider a construct to be the specific definition of an ability that provides the basis for a given assessment or assessment task and for interpreting scores derived from this task. The construct definition for a particular assessment situation becomes the basis for the kinds of interpretations we can make from the assessment performance (p.43).

Según los autores, el constructo de un examen se puede definir desde diferentes aristas; por ejemplo, a través del análisis del contenido de un curso, del análisis de necesidades de las competencias comunicativas que son necesarias para desempeñar tareas de uso de la lengua en el “mundo real” o también a partir de modelos teóricos. Sin embargo, debido a que el

constructo en la evaluación de lenguas hace alusión a aquellas habilidades de la lengua que no son directamente observables, Hamp-Lyons (2002) argumenta que éste resulta difícil de medir, pero que se puede evidenciar mediante comportamientos que representen el constructo en cuestión, y en el caso de PE, que se evalúen mediante rúbricas. En el caso particular de la escritura, Weigle (2002) comenta que debido a que el constructo es abstracto, es necesario tomar en consideración el propósito del examen y el uso de la lengua en el ámbito para el cual deseamos evaluar.

En el caso del EXELEAA, el constructo de la PE partió de la necesidad de medir las habilidades de escritura que deben poseer los estudiantes no hispanohablantes que desean ingresar a la UNAM. No obstante, debido a que la UNAM no cuenta con un documento detallado de dichas habilidades, el constructo del examen se sustentó a partir de dos fundamentos teóricos: la competencia comunicativa del modelo de Canale y Swain (1980) y los procesos de escritura desde la psicolingüística cognitiva-discursiva de van Dijk (1985) y van Dijk y Kintsch (1983). Para poder diseñar los descriptores y la rúbrica o escala de evaluación, también se tomaron en cuenta los lineamientos del Marco Común Europeo de Referencia (MCER). En el MCER se establecen los niveles de actuación en segundas lenguas, desde sus estadios iniciales hasta los más avanzados. De los seis niveles de actuación plasmados en dicho documento, se decidió optar por los bloques intermedios altos y superiores de éste, por lo cual se incluyeron los descriptores de los niveles B1+, B2 y C1 dentro de los descriptores y las tablas de especificación del examen.

1.5.1. EL SUSTENTO TEÓRICO DE LA PE DEL EXELEAA

El primer pilar sobre el cual se sustentó la PE del EXELEAA partió del concepto competencia comunicativa. En 1972, Hymes planteó cuatro características necesarias para la competencia comunicativa: lo posible, lo viable, lo apropiado y lo realizado (p.281). Sin embargo, la comunicación no implica entidades independientes, sino la interrelación entre los diferentes componentes de la lengua. Más adelante, Canale y Swain (1980) y Canale (1983) trataron de clarificar los cuatro conceptos que Hymes (1972) había definido dentro de la competencia comunicativa y que no quedaban definidos. Dentro de su marco teórico, ellos incluyeron tres aspectos importantes a considerar: la competencia gramatical, sociolingüística y estratégica. Dentro de su modelo, dichos autores consideran que la competencia comunicativa debe incluir la interacción, la creatividad, el contexto discursivo y sociocultural, el uso auténtico

de la lengua, y ser llevado y evaluado dentro de los límites del desempeño de la persona (Canale y Swain, 1980, p. 32).

Canale y Swain (1980) consideran que la competencia gramatical incluye el conocimiento léxico, las reglas de la morfología, sintaxis, semántica y fonología, mientras que la competencia sociolingüística está conformada por dos partes: las reglas socioculturales y discursivas. Las primeras, vistas desde la óptica de Hymes (1972), especifican la forma en que uno debe comunicarse apropiadamente; desde luego, partiendo de un contexto determinado. Finalmente, la competencia estratégica implica la comunicación verbal y no verbal que el individuo emplea para compensar la ruptura dentro de la comunicación.

En realidad, siguiendo el concepto chomskiano de competencia, en cuanto a lo que es posible gramaticalmente, se puede establecer una relación entre los conceptos de Canale y Swain (1980) y los de Hymes (1972) en el primer rubro. En este sentido, la competencia gramatical se encuentra íntimamente relacionada con lo posible. La otra relación entre los modelos de estos autores se encuentra en el plano de la sociolingüística, en donde ésta se asemeja a lo viable. La competencia sociolingüística, entonces, se ocupa de la coherencia discursiva. En cuanto a la competencia estratégica, Canale y Swain la definen como:

[...] This component will be made up of verbal and non-verbal communication strategies that may be called into action to compensate for breakdowns in communication due to performance variables or to insufficient competence (Canale y Swain, 1980, p. 30).

En general, la crítica que se puede hacer a los modelos comunicativos de Hymes (1972) y Canale y Swain (1980) es que representan entidades estáticas con nula interrelación entre ellos, pero que tratan de describir un acto comunicativo que es dinámico y en el que, además, todas las partes se encuentran íntimamente relacionadas. La problemática de estos modelos es que ponen en duda que la competencia comunicativa pueda servir de marco pedagógico operacional para la evaluación de lenguas (Widdowson, 2001). Según este autor, de la competencia comunicativa, lo único que podemos evaluar es la competencia lingüística inserta en un contexto pragmático y social. A partir de estas apreciaciones han surgido grandes debates sobre cómo y qué evaluar en los exámenes de lengua. Por ejemplo, surge el cuestionamiento si se debería evaluar la gramática de forma explícita o en tareas orales y escritas o si se deberían evaluar las tareas de producción oral y escrita de forma aislada o integrada.

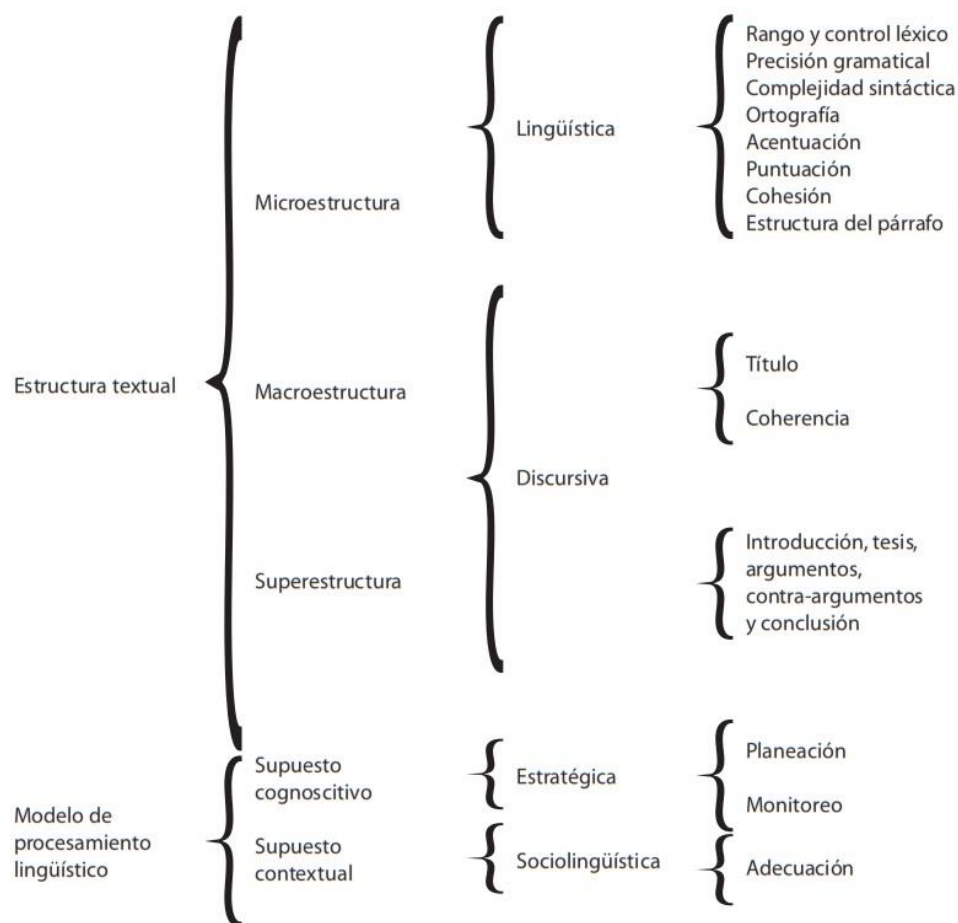
El segundo pilar sobre el cual se sustentó la PE partió de la teoría psicosociolingüística que enfoca el proceso de escritura desde una perspectiva cognitivo-discursiva que contempla la habilidad de escribir “como una compleja actividad cultural, social, cognoscitiva e interactiva” (Jurado, 2014, p. 165), que en el caso del ámbito académico se trata de la construcción de significados y la generación del conocimiento. Siguiendo la propuesta textual de van Dijk (1985), y van Dijk y Kintsch (1978) el texto se concibe en tres niveles: microestructura (en donde aplican las reglas de ortografía, puntuación y gramática), macroestructura (en donde las relaciones semánticas globales del contenido del texto se dan gracias a la coherencia y cohesión textual) y superestructura (que caracteriza al tipo de texto y permite su correcta interpretación dentro de un ámbito determinado).

1.5.2. LA EVALUACIÓN DE LA PE DEL EXELEAA

La rúbrica para evaluar la PE del EXELEAA se diseñó a partir del modelo comunicativo de Canale y Swain, la teoría psicosociolingüística de van Dijk (1985) y los lineamientos establecidos por el Marco Común Europea de Referencia para los niveles B1, B2 y C1⁷. La Figura 1.1 plasma la estructura textual y el modelo del procesamiento lingüístico-cognoscitivo.

Después de varias revisiones a la rúbrica, ésta quedó segmentada de la siguiente manera: competencia lingüística (vocabulario, precisión gramatical, complejidad sintáctica, ortografía y puntuación), competencia discursiva (desarrollo de ideas –para la tarea 1 y 2–, coherencia y cohesión), competencia pragmática (adecuación) y competencia estratégica (planeación y monitoreo del texto). El Anexo 2.2 muestra la primera versión de la rúbrica que se empujó durante el pilotaje de la presente investigación. Para cada uno de los rubros (excepto desarrollo de ideas), se asignó un rango de cero a tres puntos o de uno a tres puntos, y una puntuación final global de cero a tres puntos. En el caso del desarrollo de ideas, se tomaron valores binarios de cero y uno. Las dos tareas, en conjunto, suman un total máximo de 70 puntos. Como se mencionó anteriormente, la ponderación de la PE equivale al 20 por ciento del total del examen.

⁷ Consúltese Jurado (2014) para mayor información respecto del proceso de construcción de una rúbrica analítica basada en la teoría.

Figura 1.1: Componentes del marco teórico organizados en la rúbrica (Jurado, 2014, p. 176)

1.6 JUSTIFICACIÓN

El estudio que aquí se presenta surgió de la necesidad de contar con instrumentos y procedimientos detallados que permitieran validar un examen. En este caso, debido a lo amplio que resultaría un estudio para validar todo el examen, se decidió acotar exclusivamente la PE. Una de las razones fundamentales para elegir esta prueba se debió a que el examen que anteriormente se empleaba con fines académicos de ingreso al posgrado (EPLA) no contaba con dicha prueba. Adicionalmente, cabe destacar que la escritura es una de las habilidades primordiales a través de las cuales se evalúa el desempeño académico de los estudiantes de posgrado. Más aún, dependiendo del posgrado, el proceso de titulación para los estudiantes de posgrado es mediante la redacción de una tesis y/o la publicación de artículos de investigación en español o en inglés.

Cuando se diseñan exámenes, cuyos resultados inciden decisivamente de forma benéfica o perniciosa la vida de los sustentantes, es imprescindible contar con estudios que garanticen su validez. Por esta razón, en el caso de la presente investigación, se diseñaron estudios que permitieran evidenciar si la PE del EXELEAA mide eficazmente las habilidades de escritura que requiere un estudiante universitario o de posgrado para ser académicamente competente. De ahí que, el proceso de validación del examen resultara necesario para garantizar no sólo su relevancia, sino también su validez.

1.7 OBJETIVOS DE LA INVESTIGACIÓN

El fin de la PE del EXELEAA consiste en evaluar las habilidades de escritura de los aspirantes de posgrado no hispanohablantes que desean cursar estudios universitarios en la UNAM. Por esta razón, el objetivo general de la presente investigación consiste en validar la PE del examen. Dicho proceso de validación se encuentra circunscrito dentro del enfoque de argumentación propuesto por Kane (2013) para validar exámenes de alto impacto.

Objetivo general:

- **Investigar si la PE del EXELEAA mide eficazmente las habilidades de escritura académica que requieren los estudiantes no hispanohablantes para ser competentes a nivel posgrado.**

Debido a que la validación de una prueba consta de diversos pasos que se deben respaldar mediante evidencia, a continuación, se enuncian los objetivos particulares de acuerdo con las inferencias a las que se desea llegar.

Objetivos particulares:

- Conocer cuál es el uso y la importancia de la lengua escrita a nivel posgrado.
- Investigar qué habilidades de escritura son importantes para que un estudiante de posgrado sea académicamente competente.
- Argumentar la pertinencia de las tareas incluidas en la PE del EXELEAA.
- Determinar la calidad de la rúbrica analítica.
- Indagar si la rúbrica analítica mide eficazmente las habilidades de escritura que se requieren en el posgrado.
- Evaluar la calidad de los puntajes otorgados por los evaluadores de la PE.

- Investigar si es posible extrapolar los resultados de la PE al contexto de escritura académica de nivel posgrado.
- Conocer cuál es el proceso de escritura académica de los estudiantes de posgrado no hispanohablantes.

1.8 APORTACIONES DE LA INVESTIGACIÓN

La presente investigación conlleva aportaciones teóricas, metodológicas y pedagógicas directamente relacionadas con la lingüística aplicada y, en particular, con la evaluación de lenguas extranjeras y el proceso de validación de exámenes.

A nivel teórico, se aborda la validez como un concepto unitario (Messick 1989) que funge como un paraguas, el cual, a través de una novedosa perspectiva sustentada en la argumentación, permite establecer inferencias respecto de la validez de un examen. Dichas inferencias, a su vez, requieren ser respaldadas mediante información que justifique la validez y fuerza del argumento.

A nivel metodológico, la presente investigación, compuesta por tres estudios de tipo mixto, permitió recolectar y relacionar información proveniente de distintas fuentes en aras de proveer el sustento necesario al argumento de validación diseñado para la PE del EXELEAA. En el primer estudio, se incursiona en un campo que ha sido poco estudiado en el contexto hispanohablante: la importancia de la escritura académica a nivel posgrado y las prácticas de evaluación docente en este ciclo universitario. En cuanto al segundo estudio, se emplea el método estadístico de Rasch para facetas múltiples con el fin de dar cuenta de la calidad del proceso de validación de la PE (i.e., la calidad de la rúbrica, la dificultad de las tareas y el índice de severidad o indulgencia de los evaluadores). Si bien estos dos primeros estudios son vastos en el contexto anglosajón, aún son incipientes en otras lenguas, como es el caso del español. Finalmente, el tercer estudio contribuye a las investigaciones que se han llevado a cabo en las últimas dos décadas para dar cuenta de cómo se pueden extrapolar los resultados de un examen al “mundo real”; es decir, al contexto para el cual se evalúa, en este caso, el académico.

Finalmente, a nivel pedagógico, los tres estudios aquí plasmados representan una guía para quienes diseñan exámenes e instrumentos de evaluación. El detalle con el cual se describe la concepción de los instrumentos, su diseño y aplicación constituyen un apoyo para

quienes se encuentran en calidad de construir instrumentos para evaluar el desempeño de candidatos y evaluadores en pruebas de respuesta construida, por un lado, pero también representan una guía para quien desee llevar a cabo estudios de tipo mixto en donde convergen múltiples instrumentos y participantes.

1.9 ORGANIZACIÓN DE LA TESIS

La tesis se conforma de siete capítulos. En el presente capítulo, introductorio, se describió la importancia de la evaluación con fines académicos, así como la creciente necesidad de evaluar el español con estos fines. Más adelante, se explicó cómo se concibió la PE del EXELEAA y también la escala que se diseñó para calificar los textos que producen los examinados. Por último, se presentó la justificación, los objetivos y las aportaciones de la investigación.

En el Capítulo 2 se hace una breve reseña histórica de la validez dentro de los exámenes de lenguas extranjeras y se describe el modelo argumentativo de Toulmin. Posteriormente, se describen aquellos enfoques argumentativos que, siguiendo el modelo de Toulmin, se han diseñado en las últimas dos décadas con el fin de proveer una directriz que permita validar exámenes. Estos modelos resultan importantes para generar argumentos que, mediante evidencia, permitan establecer inferencias con respecto al desempeño de los candidatos.

En el Capítulo 3 se presenta el argumento de validación que se elaboró para validar la PE del EXELEAA. Con el fin de poder respaldar las cuatro inferencias que se establecieron para la PE, se diseñaron tres estudios de tipo mixto convergente. A su vez, cada uno de estos estudios se encuentra conformado por distintas investigaciones cuantitativas y cualitativas diseñadas específicamente para reunir evidencia suficiente que permita respaldar las inferencias del argumento de validación de la PE.

El Capítulo 4 contiene el Estudio 1, el cual se enfoca en la inferencia de la descripción del dominio y la pertinencia de las tareas incluidas en la PE. El Capítulo 5 muestra el Estudio 2, mismo que tuvo como objetivo evidenciar la calidad del proceso de corrección de la prueba escrita. Éste a su vez, da cuenta de dos inferencias: evaluación y generalización. La inferencia de evaluación busca respaldar la calidad del proceso de corrección a través de su rúbrica, mientras que el de generalización pretende dar cuenta de la calidad y consistencia de la evaluación por parte de los correctores. Finalmente, el Capítulo 6 presenta el Estudio 3, el cual

incluye la inferencia de extrapolación. Esta inferencia pretende evidenciar hasta qué punto se pueden extender los resultados de un examen al uso y práctica de las habilidades de escritura requeridas en el contexto académico.

El Capítulo 7 contiene discusiones generales respecto de los hallazgos de los tres estudios anteriormente descritos, así como de la implementación del enfoque de la argumentación para validar la prueba escrita del examen de español con fines académicos. Finalmente, se ofrecen conclusiones generales también en cuanto a los estudios y el argumento de validación de la prueba.

Al final de la tesis figuran las referencias y los anexos, mismos que se encuentran segmentados de acuerdo con los tres estudios descritos. A continuación, en el Capítulo 2, se presentarán la discusión sobre los distintos modelos y enfoques de validación de exámenes basados en la argumentación.

CAPÍTULO 2

LA VALIDEZ EN LOS EXÁMENES DE LENGUAS EXTRANJERAS

2.1. INTRODUCCIÓN

La validez en la evaluación de los exámenes de lenguas de alto impacto representa un desafío para las instituciones que se dedican a diseñar exámenes. Debido a que estos exámenes juegan un papel primordial en la vida de los examinados, es imprescindible que se evalúe de manera justa y equitativa a los candidatos que cada año sustentan dichos exámenes. Por tal motivo, la ética en la evaluación ha adquirido un lugar tan importante como la validez y confiabilidad de la misma. Sin embargo, una de las dificultades que ha girado en torno a la noción de validez se debe a que los estándares internacionales (APA, AERA y NCME¹, 1974, 1985, 1999) conciben la validez como una serie de elementos disgregados y desarticulados que resultan difíciles de incorporar al momento de diseñar y validar exámenes. Por tal motivo, desde comienzos del siglo del siglo XXI, la validez se ha concebido como un argumento que requiere de ciertas inferencias, evidencias y sustento que respalde la interpretación de los resultados y la toma de decisiones.

En este apartado teórico, se hace una breve reseña de los orígenes de la noción de validez en los exámenes, partiendo de los conceptos aportados por Cronbach (1988) y Messick (1989), así como con el esfuerzo por incluirlos dentro de los modelos de diseño y de validación en los exámenes durante la década de los 90. Más adelante, se describe la problemática que se suscitó a finales del siglo XX al tratar de poner en práctica los modelos teóricos para validar exámenes y los conceptos de validez fragmentados que eran utilizados por los expertos en evaluar exámenes de alto impacto. Posteriormente, se describe el modelo de argumentación de Toulmin (1958) y la aplicación que ha tenido en la evaluación a través de quienes lo han puesto en práctica para la generación de diversos enfoques de validación de exámenes.

¹ APA: *American Psychological Association*
AERA: *American Educational Research Association*
NCME: *National Council on Measurement in Education*

2.2. LOS ORÍGENES DE LA VALIDEZ EN LOS EXÁMENES

La validez de cualquier examen de lengua de alto impacto es indispensable para tomar decisiones justas y equitativas con base en información veraz respecto de las habilidades lingüísticas que posee un sustentante. En un examen, la validez se desarrolla previa, durante y posteriormente a su administración, y digamos que su fin, en términos generales, es indagar si un examen en realidad mide lo que debería medir. Sin embargo, la manera de efectuar dicha validez ha variado sustancialmente desde su florecimiento en la década de los 50 del siglo XX hasta nuestros días.

Durante la primera mitad del siglo XX, el concepto de validez se sustentaba en modelos estadísticos basados en el criterio de los exámenes; es decir, la validez era definida como la correlación entre el puntaje de un examen y el puntaje definido como criterio. Este criterio, por lo general, se encontraba dado por exámenes similares que evaluaban las mismas habilidades; por lo tanto, un examen exclusivamente se consideraba como válido si media lo mismo que otros instrumentos ya existentes.

Cronbach y Meehl (1955) introdujeron cuatro tipos de validez: predictiva, concurrente, de contenido y de constructo, las cuales constituyeron los ejes fundamentales de la validez de exámenes hasta finales del siglo XX. Dentro de estas categorías, destaca la validez de constructo, la cual, según los autores, “se ve involucrada cuando un examen ha de ser interpretado como una medida de algún atributo o cualidad que no es *definida operacionalmente*” (p. 282). Esta idea de constructo partió de la necesidad de concebir la validez en términos científicos, en la cual el constructo es definido por la teoría y, por lo tanto, la interpretación del constructo puede ser evaluada empíricamente junto con la teoría que lo sustenta. Previo a la noción de la validez de constructo, la interpretación de los resultados de un examen era considerada como algo simple, claro y evidente; no obstante, a partir de la década de los 50, la interpretación de un constructo teórico involucró un riguroso análisis científico.

En 1974, los organismos estadounidenses encargados de desarrollar los estándares a considerar en la elaboración de exámenes, APA, AERA y NCME, definieron los cuatro tipos de validez descritos por Cronbach y Meehl (1955): predictiva, concurrente, de contenido y de constructo. Desafortunadamente, la problemática que se suscitó a finales de esta década fue que los diferentes métodos de validez eran tratados como diversos instrumentos que podían ser empleados en distintas situaciones de acuerdo a cómo los evaluadores consideraban que fuera

más conveniente emplearlas. El punto fundamental era que no se contaba con un marco general de validación para los exámenes, puesto que la validez de contenido y concurrente dependían del tipo de examen que se quería validar. Consecuentemente, debido a que la validez de constructo era la única que contaba con un sustento científico, se adoptó como marco para la validación de exámenes (Messick, 1989).

Pese a que la validez de constructo propuesta por Cronbach y Meehl (1955) causaba dificultades en su aplicación, uno de los mayores logros de los autores fue demostrar que la interpretación de los resultados no podía darse por hecho y que se necesitaba de una serie de afirmaciones circunscritas dentro un marco teórico. En realidad, una de las grandes aportaciones de Cronbach y Meehl al concepto de validez se debió a la concepción de ésta no como un mecanismo matemático, como es el caso de la confiabilidad, sino más bien como todo un proceso denominado validez de argumento, posteriormente retomado por Kane (1992) en su enfoque de validación de exámenes, el cual consiste en recabar evidencia para establecer inferencias, más que exclusivamente para validar instrumentos. Gracias al trabajo de Cronbach a lo largo de varias décadas, se cimentaron las bases que propiciaron el surgimiento del modelo de validez de Samuel Messick en 1989. Este modelo de validez, aún vigente, es el más representativo y sobre el cual versan las discusiones que hoy en día existen en materia de validez en la evaluación de exámenes de alto impacto. Messick (1989) concibió la “validez como un juicio evaluativo integrado del grado en el cual la evidencia empírica y las justificaciones teóricas apoyan la adecuación y la pertinencia de las inferencias y las acciones basadas en los resultados de los exámenes o de otros modos de evaluación” (p. 13).

En este sentido, la mayor aportación de Messick fue concebir la validez como un aspecto unitario y no como una serie de elementos disgregados y desarticulados. Como se puede apreciar en el Cuadro 2.1, desde la óptica de Messick, la evaluación es un procedimiento para extraer inferencias (toma de decisiones) a partir de ciertas evidencias (exámenes); de tal suerte que el objetivo primordial de la validez empírica de los exámenes sea demostrar que las interpretaciones y toma de decisiones respecto al desempeño de un sustentante en un examen de dominio son justas y equitativas. En suma, una de las grandes aportaciones de Cronbach y Messick fue la concepción de validez como un proceso de argumentación, misma que tuvo gran influencia en los modelos de validez de la primera década del siglo XXI.

Cuadro 2.1: Facetas de validez como una matriz progresiva (Messick 1995, p. 748)

(Traducido por el investigador)

	INTERPRETACIÓN DEL EXAMEN	USO DEL EXAMEN
BASE PROBATORIA	Validez de Constructo (VC)	VC + Relevancia / Utilidad (R/U)
BASE CONSECUENTE	VC + Valor de las Implicaciones (VI)	VC + R/U + VI + Consecuencias sociales

2.3. LA VALIDEZ A FINALES DEL SIGLO XX Y COMIENZOS DEL SIGLO XXI

Aun cuando el modelo de Messick (1989) constituyó un parteaguas en la noción de validez al incluir las consecuencias y el uso del examen dentro de su matriz de facetas de validez, una de las mayores dificultades de este modelo se debió a su materialización; en otras palabras, a lo abstracto y a la poca claridad como guía para aplicarlo al proceso de validación de exámenes (Bachman, 2005; Kane, 1992; Mislevy, Steinberg y Almond, 2003). Por tal razón, a finales del siglo XX y comienzos del XXI, la directriz para validar exámenes consistía más bien en una serie de características necesarias dentro de la validez, pero que no constituían un procedimiento lógico mediante el cual se encontraban vinculadas y relacionadas unas de otras (Bachman, 2005). Pese a las dificultades para materializar el modelo de Messick, la inclusión del significado social, los valores, la ética y las consecuencias de la evaluación fomentaron la necesidad de incorporarlos dentro del diseño y validación de exámenes.

Por ejemplo, en el modelo de validación de exámenes de lengua, Bachman y Palmer (1996) describieron seis cualidades que debían ser consideradas dentro de la utilidad de un examen; a saber: confiabilidad, validez de constructo, autenticidad, interactividad, impacto y practicidad. Paralelamente, otros autores como Lynch (2001) y Kunnan (2003) discutieron nociones de ética y de imparcialidad en la validez y toma de decisiones. No obstante, también de manera fragmentada, todos estos autores generaron listas de cualidades que debían poseer los exámenes para ser válidos, justos y equitativos. Lynch describió un modelo de validez en el cual integró consideraciones éticas bajo cinco categorías: imparcialidad, autenticidad ontológica, la autenticidad a través de la referencia, validez de impacto y de consecuencia, y la evolución de las relaciones de poder. Kunnan, por su parte, también elaboró un modelo de validez denominado

“evaluación imparcial”, en el cual incorporó las consecuencias de los exámenes y la toma de decisiones (Bachman, 2005).

Sin embargo, como afirma Bachman (2005), todas estas características son importantes dentro de la concepción y operacionalización de un examen, mas constituían solamente una lista con cualidades que sí corresponden a la validez, aunque no dan cuenta de qué forma se vinculan ni se interrelacionan para garantizar el correcto uso de los exámenes en la toma de decisiones y, por lo tanto, sus consecuencias:

... what we have at present, as a basis for justifying test use, are essentially lists of qualities and questions that test developers and test users need to consider, with no clear logical mechanism for integrating these into a set of procedures for test developers and users to follow (Bachman, 2005, p. 7).

Por estos motivos, a finales del siglo XX, debido a la falta de integración e interrelación entre los diversos conceptos de validez, el derrotero de la validez dentro del ámbito de lenguas comenzó a orientarse hacia la incorporación de un argumento que fungiera como “paraguas” y que no sólo describiera los componentes de validez, sino que además los integrara e interrelacionara entre sí. De ahí que el modelo de argumentación de Toulmin (1958) fuera retomado por los teóricos que a lo largo de dos décadas han incorporado dicho modelo con el fin de concebir la validez de forma más global que fragmentaria. Por lo tanto, antes de describir los modelos más influyentes en la validez de exámenes de lengua, será necesario describir el modelo de argumentación de Toulmin.

2.4. EL MODELO DE ARGUMENTACIÓN DE TOULMIN (1958) Y TOULMIN, RIEKE Y JANIK (1984)

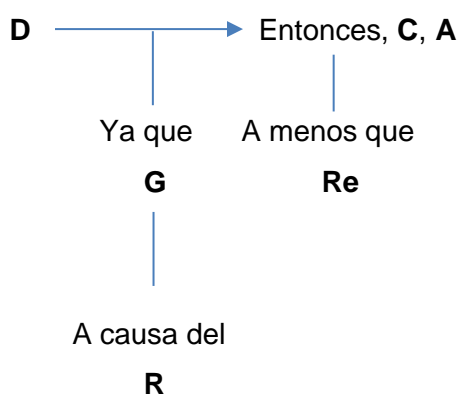
El modelo de argumentación propuesto por Toulmin (1958), y revisado y ampliado por Toulmin, Rieke y Janik (1984), ha tenido aplicaciones en diversos ámbitos como el derecho, las ciencias naturales, la medicina y la estética; no obstante, no fue sino hasta la década de los 90 cuando comenzó a vislumbrarse su adecuación dentro del proceso de validación de exámenes (Kane, 1992). En la Figura 2.1 se puede apreciar el modelo de argumentación de Toulmin. A continuación, se describirá con detalle cada uno de los componentes de dicho modelo.

Toulmin (1958) considera como punto de partida en el proceso de argumentación dos componentes: una aserción (*claim*) y los hechos o datos (*data*) que fungen como base o

fundamento para validar dicha aserción. Sin embargo, el verdadero reto, afirma el autor, no se encuentra representado por la aserción que se establece, sino por el procedimiento seguido para llegar a ella. Partiendo de esta premisa, el autor menciona que lo importante no es fortalecer nuestros datos, sino las proposiciones que de ellos podamos derivar: reglas, principios o inferencias. A este tipo de inferencias las denomina garantías (*warrants*); las cuales, aunque puedan parecer obviedades, en realidad no lo son.

Figura 2.1: Modelo de argumentación de Toulmin² (1958, p. 104)

(Traducido por el investigador)



D – Datos
 G – Garantías
 R – Respaldo
 C – Calificadores modales
 Re – Refutación
 A – Aserción

² Toulmin (1958, p. 105) proporciona el siguiente ejemplo con base en su modelo: Harry nació en Bermudas (datos); por lo tanto, es británico (aserción), debido a que los hombres que nacen en Bermudas son británicos (garantías) con base en los estatutos y disposiciones legales (fundamentos). Lo anterior sería contundentemente cierto, siempre y cuando (refutación) los padres de Harry fueran extranjeros o él se hubiera naturalizado estadounidense, por ejemplo.

El autor proporciona este ejemplo para distinguir los datos o fundamentos de las garantías, pues las primeras son explícitas; mientras que las segundas, implícitas. Las garantías no se validan por sí mismas, razón por la cual es necesario que exista respaldo detrás de ellas. En el ejemplo anterior, para poder hacer una aseveración tal como: "Los hombres nacidos en Bermudas son británicos", debemos respaldar dicha afirmación, ya que como afirma Toulmin, la mayor dificultad dentro del proceso de argumentación yace precisamente en especificar las garantías y fundamentar la naturaleza de su "autoridad". Como se mencionó anteriormente, las garantías son presuposiciones, las cuales deben ser justificadas o respaldadas. El respaldo debe ser considerado cuidadosamente, puesto que su pertinencia se ubica precisamente en el campo o ámbito en el cual el argumento es enunciado.

El problema de las garantías, menciona Toulmin, es que en ocasiones existen varias formas de argumentar en distintas direcciones, cada una con sus respectivas premisas y evidencias. Ante estas circunstancias, se deben sopesar los distintos argumentos en conflicto, de tal suerte que se pueda elegir aquella línea de argumentación que conlleve mayor peso. Ahora bien, las garantías son presuposiciones que no se validan per se, sino mediante el respaldo (*backing*) detrás de ellas que permite justificarlas. Este respaldo debe ser considerado cuidadosamente, puesto que su pertinencia se ubica precisamente en el campo o ámbito en el cual el argumento es enunciado. Los cuatro puntos mencionados anteriormente, agrega Toulmin, son indispensables para conferirle solidez al argumento.

Una vez que el argumento cuenta con solidez, el siguiente paso consiste en establecer la fuerza del mismo. Toulmin et al. (1984) comentan que los cuatro pasos anteriores son sólo el principio del argumento, puesto que las garantías son de diversa naturaleza, por lo cual pueden conferir diferentes grados de fuerza de conclusión. Es aquí donde entra el segundo nivel de análisis, mismo que los autores denominan como la fuerza de un argumento, la cual se encuentra compuesta por los calificadores modales (*qualifiers*), refutación (*rebuttal*) y excepciones (*exceptions*). Según Toulmin (1958) y Toulmin et al. (1984), esta fuerza se expresa mediante adverbios que denotan una cualidad: *necesariamente*, *tentativamente*, *probablemente*, *aparentemente*, *quizás*, *muy probablemente*, *seguramente*, etc. Por ende, puede que no resulte suficiente especificar los datos, la aserción y la garantía, ya que quizás sea necesario añadir un calificador modal. El calificador se encuentra determinado y complementado por las refutaciones y/o por las condiciones de excepción, y se localiza justo frente a la conclusión, puesto que determina las condiciones mediante las cuales se puede aseverar algo.

Por último, las refutaciones y excepciones ayudan a establecer los límites de la aserción mediante un calificador modal; asimismo, permiten afirmar o rechazar un argumento. Toulmin et al. (1984) clasifican las reservas en dos tipos: a) cuando los datos, las garantías y el respaldo dan apoyo a la aserción solamente de forma parcial y b) cuando existe soporte de los datos, garantías y respaldo solamente en ciertas condiciones. En el primer caso, el argumento carecerá de toda su fuerza, mientras que en el segundo, el argumento adquiere validez en la ausencia de condiciones particularmente excepcionales.

Sin lugar a dudas, gracias a la adopción del modelo de argumentación de Toulmin en la validación de los exámenes de lenguas, se dio un cambio a la noción de validez y confiabilidad.

Asimismo, otra de las ventajas del modelo de Toulmin se debe a que éste no busca evidencia perfecta, sino más bien plausible.

If we always waited until absolutely rigorous arguments could be constructed before we acted with reasoned confidence, we would be overtaken by events before we had occasion to act. In practice, therefore, it is often reasonable to base our conclusions on something less than absolutely perfect evidence. We then put forward our claims, not as being formally irrefutable, but rather as being practically strong or reliable (Toulmin et al. 1984, p. 81).

Esta noción de verosimilitud versus perfección en la evidencia resulta indispensable en la construcción de un argumento de validación, puesto que, la suma de los elementos de validez no necesariamente le confiere validez a un examen, sino más bien la validez se vale de diversos mecanismos para ser plausible. Aunado a lo anterior, el argumento empleado puede irse modificando a lo largo del tiempo en caso de que nuevos hallazgos fundamenten la implementación de cambios en el modelo previamente concebido. Más aún, el modelo de Toulmin (1958) no solamente se modifica a lo largo del tiempo, sino también de manera sincrónica, pues existen numerosas perspectivas de cómo abordar la evaluación de las competencias comunicativas de los sustentantes. Por ejemplo, simplemente en el caso de las PE, en ocasiones se emplean pruebas indirectas para inferir las habilidades de escritura de un examinando, mientras que en otras, se emplean pruebas directas que podrán ser revisadas por un corrector humano, dos o incluso un corrector electrónico. Así pues, cada uno de los caminos que se deseen adoptar para evaluar requerirán de un argumento diferente y, por lo tanto, de inferencias y evidencias acordes con éste.

En el caso de la evaluación de lenguas, por ejemplo, ésta adquirió rigor científico y metodológico a partir de la publicación en 1961 del libro *Language Testing* de Robert Lado. A partir de este momento, el desarrollo de la evaluación de lenguas dio un giro de un rol periférico en la lingüística aplicada a una posición central dentro de esta rama. Esto fue posible gracias a que el estatus de la evaluación pasó de ser un instrumento práctico a uno también teórico, el cual cuenta con poder explicativo de fenómenos relacionados con el aprendizaje de lenguas. En *Language Testing*, Lado incorporó el conocimiento de la lingüística moderna a los exámenes de lengua. Desde entonces, la visión de la evaluación de lenguas ha pasado por diversas fases hasta la reciente incorporación del argumento de validación para diseñar y validar exámenes.

2.5. LOS ENFOQUES DE ARGUMENTACIÓN APLICADOS A LA EVALUACIÓN DE EXÁMENES

Actualmente, tras la aportación de Messick (1989) al concebir la evaluación como un concepto unitario, la validez de un examen dejó de restringirse solamente a métodos psicométricos y dio lugar a un nuevo paradigma de validez basado en la argumentación. Este tipo de apertura ha permitido que la validez se conciba a partir de un espectro más amplio que incluya múltiples tipos de evidencia cualitativa y cuantitativa que den respaldo a un examen.

A continuación, se presentan los tres modelos/enfoques de argumentación dentro del proceso de validación que mayor influencia han tenido en los exámenes de lenguas. A lo largo de dos décadas, éstos se han ido desarrollando y retroalimentando entre sí; sin embargo, no fue sino hasta comienzos del siglo XXI que éstos comenzaron a adquirir mayor fuerza y consistencia. Por esta razón, hace apenas una década que dichos modelos/enfoques comenzaron a ser adoptados dentro del proceso de validación de exámenes de lengua. El orden en el cual se presentan corresponde meramente a un criterio alfabético; posteriormente, se discutirán las aportaciones de cada uno de ellos al campo de la evaluación de lenguas.

2.5.1. EL ENFOQUE BASADO EN EL ARGUMENTO DE VALIDACIÓN DE BACHMAN

Bachman (2005) reconoce la importancia que juega la validez en relación con las consecuencias del uso de exámenes de lenguas y enfatiza la necesidad de contar con un enfoque argumentativo que permita llegar a conclusiones o afirmaciones respecto de una forma de evaluación específica, un grupo de examinandos específico y un contexto específico.

Tras poco más de dos décadas de la publicación de Bachman: *Fundamental considerations in language testing*, en 1990, la discusión en torno a la problemática de la evaluación continúa vigente. Después de una década de discusión respecto al tema, Bachman (2005) retoma dos preguntas fundamentales formuladas por Spolsky (1981) en las primeras discusiones en materia de evaluación a principios de la década de los 80: “*How sure are you of your decisión?*” y “*How sure are you of the evidence that you’re using to make that decisión?*” (Spolsky, 1981 citado en Bachman, 2005, p. 5).

Ambas preguntas denotaban ya la importancia que representaba la toma de decisiones justa y equitativa. Bachman (2005, p. 5) retoma ambas preguntas y las reformula de la siguiente manera: “*How convincing is the argument for using the assessment in this way?*” y la segunda,

“How credible is the evidence that supports this argument?” Para Bachman, un aspecto fundamental en su trabajo de validación se ha referido precisamente al propósito y al uso que se le dará a los exámenes. El enfoque de Bachman (2005) retoma la estructura de argumento de validación elaborada por Mislevy et al. (2003), la cual se sustenta en el modelo de argumentación de Toulmin. Bachman, enfocando de manera práctica el modelo de Toulmin y relacionándolo con la evaluación de lenguas, describe sus partes de la siguiente manera:

- **Aserción:** Se refiere directamente a la interpretación que queremos establecer. Es una conclusión sobre lo que el examinando conoce y puede hacer.
- **Datos:** Es la información en la cual se sustenta la aserción; es decir, en las respuestas del examinando.
- **Garantía:** Son las proposiciones que empleamos para justificar las inferencias.
- **Respaldo:** En la evaluación, ésta proviene de la teoría, experiencia previa o evidencia del proceso de validación, así como del análisis de necesidades de la tarea del dominio que se desea evaluar mediante el examen.
- **Refutación:** Corresponde a las posibles fuentes de invalidación.
- **Evidencia de refutación:** Consiste en la evidencia que apoye, debilite o rechace una explicación alternativa (2005, p. 9-10).

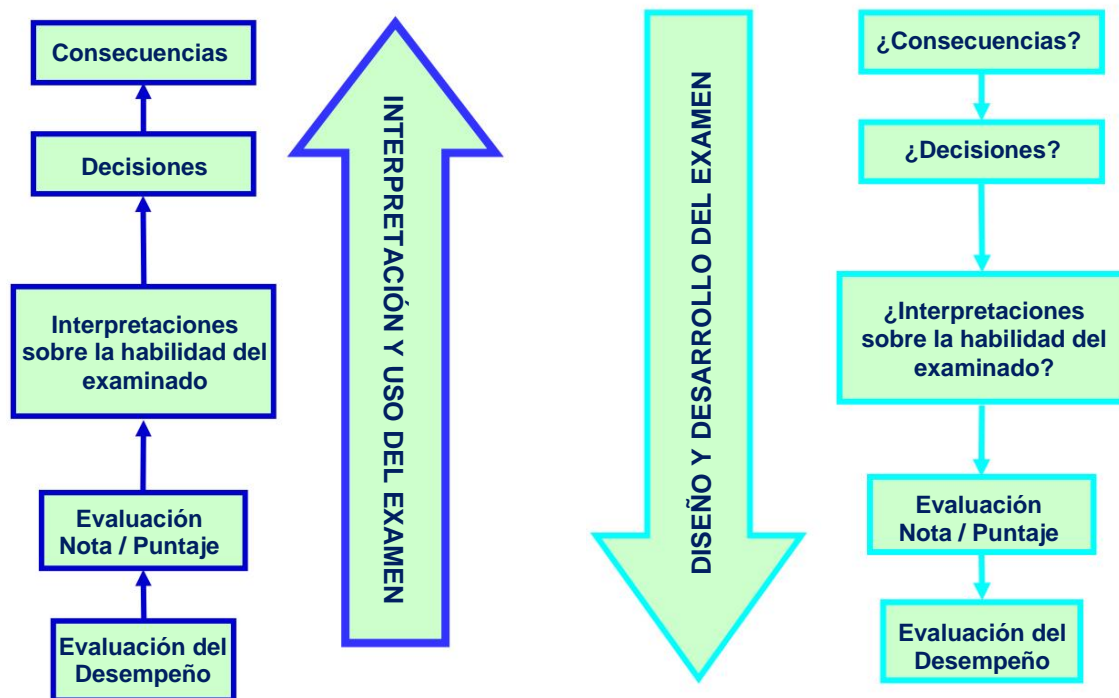
Bachman y Palmer (2010) retoman nuevamente las premisas básicas del enfoque argumentativo de Bachman (2005), pero parten de la equidad en el uso de exámenes; es decir, en las consecuencias del uso de la evaluación por las partes interesadas, que generalmente son quienes están en posición de tomar decisiones sobre la admisión o rechazo de un candidato. En consecuencia, la responsabilidad en la toma de decisiones resulta un punto crucial en la validez y la equidad en los exámenes; de tal suerte que quienes los diseñan requieren de herramientas que les permitan rendir cuentas a las partes interesadas sobre la manera mediante la cual basan sus conclusiones respecto del desempeño de un candidato.

Una decisión errónea, tanto en perjuicio como en beneficio de un examinando, no solamente afectará la noción de equidad, sino que además mermará la confianza por parte de los examinandos y de las instituciones. Sin embargo, dicha responsabilidad es bipartita; es decir, por una parte, quienes elaboran exámenes deben garantizar que éstos sean confiables y que la información proporcionada a las partes interesadas les permita tomar decisiones que conlleven consecuencias benéficas para: examinandos, administradores escolares, autoridades o agencias

educativas. No obstante, por otra parte, es responsabilidad también de los usuarios la manera mediante la cual emplean los resultados y, sobre todo, tomar en cuenta el propósito para el cual fueron diseñados. Para ello, Bachman (2005) y Bachman y Palmer (2010) proponen el argumento de uso de la evaluación: *Assessment Use Argument* (AUA).

La Figura 2.2 muestra el AUA que vincula el desempeño de un sustentante en un examen con la toma de decisiones y sus consecuencias, por una parte, mientras que por otra, de manera inversa, provee el fundamento y las bases para justificar las decisiones que se toman para el diseño y desarrollo de exámenes. Por consiguiente, el enfoque de Bachman y Palmer (2010) cumple dos propósitos: servir como argumento de interpretación y uso, y también como esquema para el diseño y desarrollo de exámenes.

Figura 2.2: Argumento de uso en la evaluación (Bachman y Palmer, 2010, p. 91)
(Traducido por el investigador)



En la Figura 2.2 se puede apreciar la relación que existe entre las consecuencias, las decisiones, las interpretaciones sobre la habilidad del examinado, el puntaje que obtuvo en su evaluación y el desempeño del examinado en las tareas presentadas en la evaluación. Como se puede observar en la flecha descendente de la derecha, la relación entre los elementos antes descritos provee las bases para el diseño de los exámenes de lenguas. Podemos ver tres interrogantes que deben hacerse los diseñadores de exámenes al momento de concebir un examen: ¿cuáles son las consecuencias de la evaluación?, ¿qué decisiones serán las que se tomarán con base en los resultados del examinado?, y ¿qué interpretaciones se pueden llevar a cabo con base en el desempeño del sustentante y del resultado que se le otorga en un examen de lengua? Estas preguntas son esenciales al momento de concebir un examen, a fin de conocer el impacto que conllevará en la vida de los sustentantes.

Por otra parte, en cuanto a la interpretación y al uso que se le da al examen, podemos apreciar la relación que existe en la flecha ascendente de la izquierda de la Figura 2.2. En este caso, para poder interpretar los resultados del examinado, debemos primero contar con la evaluación del desempeño en la o las tareas que se le presentan. Posteriormente, dependiendo de la tarea, se le otorga un puntaje que sirve para interpretar cuáles son las habilidades del examinado. Una vez que se interpretan los resultados, se toman decisiones. Dependiendo de la importancia de las decisiones, será el impacto de las consecuencias.

2.5.2. EL ENFOQUE BASADO EN EL ARGUMENTO DE VALIDACIÓN DE KANE

Kane (1992), siguiendo el modelo de argumentación de Toulmin, fue el primero en considerar la validación de exámenes como un proceso basado en la argumentación. Esto debido a dos razones fundamentales. En primer lugar, porque los estándares para la evaluación educativa (AERA, APA Y NCME; 1985, 1999) no ofrecían realmente una guía apropiada para establecer el vínculo entre los resultados de los exámenes y la interpretación para hacer uso de ellos eficazmente y de manera imparcial. En segundo lugar, debido a que el marco de validez ofrecido por Messick (1989) resultaba abstracto y difícil de aplicar. En consecuencia, Kane desarrolló el enfoque basado en la argumentación concebido como “un marco sustentado en la interpretación para recolectar y presentar evidencia válida y explícita asociada con la plausibilidad de varias suposiciones e inferencias involucradas en la interpretación” (Kane, 1992, p. 2). El enfoque, desde sus inicios, se concibió compuesto por dos tipos de argumentos: el interpretativo y el de validez. El primero contenía el razonamiento que partía de los resultados de

un examen hacia las afirmaciones y decisiones que se efectuaban con base en éstos; mientras que el segundo, presentaba el caso de plausibilidad, débil o fuerte, del primero.

A lo largo de más de dos décadas, Kane (1992, 2002b, 2006, 2013) ha modificado y adaptado su enfoque de argumentación para la validación: *Argument-Based Approach to Validity*, de acuerdo a las necesidades que surgen en materia de evaluación. El autor postula que es necesario un argumento mediante el cual se sustente la relación subyacente entre el resultado de un examen y la interpretación que de éste se deriva. “Validar la interpretación del puntaje de un examen es sustentar la plausibilidad del argumento de interpretación correspondiente con su evidencia apropiada” (Kane, 1992, p. 527).

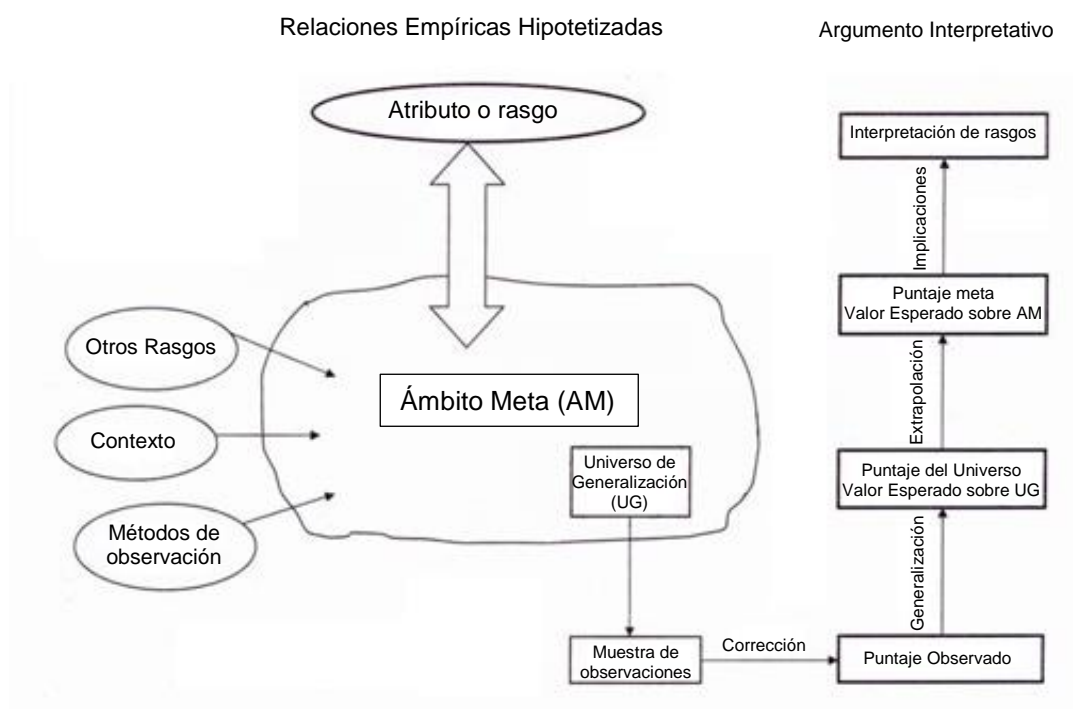
Dentro del argumento de validez, el autor enfatiza la necesidad de contar con varios tipos de evidencia organizados de manera coherente y convincente, además de su posible refutación. El proceso de argumentación puede ser debilitado mediante la evidencia, pero enmendado en sus fallas y deficiencias; sin embargo, agrega Kane (2013), el mayor problema en la evaluación se debe a suposiciones “ocultas”; es decir, aquellas que se han omitido dentro del proceso de interpretación. Evidentemente, cualquier falla dentro del argumento de validez podría ser corregida o planteada de distinta forma, pero ¿qué sucedería si se omitieran suposiciones importantes no por negligencia, sino por desconocimiento de su existencia? Por ello, Kane (2013) afirma que es importante que constantemente se evalúen los argumentos que se postulan, con el fin de que se cerciore la pertinencia de cada uno de ellos.

Ya desde su concepción inicial, Kane (1992, 2002b, 2006) y Kane, Crooks y Cohen (1999) concibieron dentro del argumento interpretativo la concatenación de una serie de inferencias que incluyen: generalizaciones, extrapolaciones, predicciones, explicaciones, inferencias basadas en la teoría y decisiones basadas en los puntajes obtenidos en una prueba. Sin embargo, si bien es cierto que a lo largo de dos décadas el modelo de Kane había girado exclusivamente en torno al argumento de interpretación de los resultados, recientemente introdujo el argumento de uso de un examen, propuesto por Bachman (2005) y retomado por Bachman y Palmer (2010) dentro de su enfoque de validación. De ahí que, Kane ahora reconozca que “la validez no es una propiedad del examen, sino más bien es una propiedad de las interpretaciones y usos que se proponen de los resultados de un examen” (Kane, 2013, p. 3). Así que, el procedimiento para conferirle validez a un examen, bajo el enfoque basado en la argumentación, consiste en desarrollar argumentos

de interpretación y de uso del examen (*interpretation/use argument*, IUA) y, posteriormente, validar dichos argumentos.

La Figura 2.3 muestra el esquema de estas relaciones que se establecen entre las distintas inferencias concebidas en el enfoque de argumentación de Kane. En la parte superior central se puede observar el rasgo o atributo que se desea observar. En el caso de los exámenes de lenguas, estos atributos se encuentran definidos por las habilidades que deseamos evaluar en nuestros examinados (p.ej., comprensión aditiva y de lectura, expresión oral y escrita). Estas habilidades se encuentran insertas en el ámbito meta; es decir, el contexto o situación comunicativa en la cual el sustentante debe demostrar que es comunicativamente competente.

Figura 2.3: Procedimiento de medición y argumento para la interpretación de rasgos (Kane, 2006, p. 33)³
(Traducido por el investigador)



Sin embargo, dado que en los exámenes sólo se puede evaluar una pequeña porción de las distintas situaciones que podrían presentarse en el "mundo real", se debe restringir el contenido del examen a lo que se conoce como universo de generalización. En la Figura 2.3, se

³ Es importante mencionar que este enfoque es de 2006 y este esquema no hace mención al uso y consecuencias de los exámenes; conceptos que incorpora en 2013.

puede observar el pequeño recuadro (universo de generalización) –perfectamente delimitado, como deberían ser los exámenes de lenguas– inserto en el círculo irregular (ámbito meta) –que no se encuentra perfectamente definido, pues las situaciones de comunicación son diversas y la frontera entre un ámbito y otro son difusas–. A continuación, se describirán las inferencias que figuran en los recuadros enlazados por las fechas de la Figura 2.3.

2.5.2.1. EL ARGUMENTO DE INTERPRETACIÓN Y DE USO

De manera muy general, el enfoque actual de Kane (2013), se divide en dos tipos de inferencias: semánticas y políticas. Las primeras incluyen la inferencia de los resultados observados, de generalización y de extrapolación; en tanto que las segundas, exclusivamente engloban la toma de decisiones.

2.5.2.1.1. INFERENCIA DE LA EVALUACIÓN

De acuerdo con Kane (1992, 2002b, 2006, 2013), el primer paso necesario para el proceso de argumentación de evaluación es contar con aquellas evidencias producidas por el alumno; ya sean escritas, orales, de comprensión de lectura o auditiva. Una vez que se tienen esas muestras, se las evalúa y se establece la primera inferencia, la cual consiste en afirmar que el puntaje observado es reflejo del comportamiento observado. En otras palabras, se busca demostrar que los mecanismos de evaluación fueron los apropiados y que éstos fueron aplicados de la manera correcta. Esta inferencia nos lleva del desempeño observado por un examinando al puntaje observado.

2.5.2.1.2. INFERENCIA DE GENERALIZACIÓN

El segundo tipo de inferencia se refiere a la confiabilidad. En el caso de las pruebas de respuesta seleccionada, se espera que los ítems sean representativos homogéneamente en grado de dificultad del universo de conocimientos que se desean evaluar; mientras que en relación con las pruebas de respuesta construida, se espera que las rúbricas estén bien elaboradas, que los correctores sepan emplearlas efectivamente y que, por lo tanto, su evaluación sea consistente. En realidad, la importancia del puntaje se debe a que con base en él se toman decisiones y se busca establecer o afirmar lo que el candidato será capaz de demostrar en contextos más amplios. Kane (2013) destaca que en un examen solamente se puede evaluar una pequeña muestra de desempeño y no toda una amplia gama de posibilidades existentes en un universo del ámbito para el cual es diseñado un examen.

2.5.2.1.3. INFERENCIA DE EXTRAPOLACIÓN

El tercer tipo de inferencia es el que involucra la extrapolación de los resultados obtenidos en la prueba hacia un contexto circunscrito en el “mundo real”; en este caso, el académico. Este tipo de inferencia es crucial y una de las más difíciles de considerar, pues implica extrapolar el desempeño de un candidato de un examen a las situaciones reales. La extrapolación es compleja e incluye ciertos procesos que van más allá de los análisis estadísticos. En este caso, se busca elaborar aserciones que contemplen situaciones de uso de la lengua dentro del “mundo real”. A este respecto, cabe señalar que Kane (2013) menciona que el argumento que mayor atención amerita es aquel que resulta más débil. En el caso de los exámenes de lengua, el argumento más endeble es, justamente, el de la extrapolación.

Kane (2006) menciona dos tipos de evidencia que puede emplearse para evaluar la inferencia de extrapolación: analítica y empírica. La evidencia analítica se emplea generalmente durante el diseño y constructo del examen, mientras que la empírica busca relacionar los puntajes observados en un examen con otros tipos de muestras o de puntajes asociados con el dominio. Recordemos que el fin último de los exámenes es tomar decisiones, por lo cual la inferencia de extrapolación resulta indispensable dentro del argumento de interpretación.

El mayor problema, como se describió anteriormente, radica en qué tan estandarizado se encuentre el examen, pues nuestra confianza en la inferencia de extrapolación dependerá de la relación que subyace entre el examen y el ámbito meta; de tal suerte que, aquellos exámenes que cubran un amplio espectro del dominio en cuestión podrán considerarse plausibles y, por ende, la inferencia de extrapolación no requerirá demasiado soporte empírico. Sin embargo, en caso contrario, cuando las tareas de un examen difieran substancialmente del dominio meta, el respaldo de la extrapolación será imprescindible. Si el argumento que se establece a partir de las inferencias de interpretación y de uso es validado mediante los mecanismos descritos anteriormente, poseerá una fuerte presuposición en favor de éste, con lo que se podrá justificar el uso de los puntajes obtenidos en un examen.

2.5.2.1.4. INFERENCIA DE TOMA DE DECISIONES

La cuarta y última inferencia implica la toma de decisiones. Ésta es la única inferencia que Kane (2013) considera de tipo político, dado que involucra considerar ciertas normas, previamente establecidas por la autoridad en cuestión (universidades, empresas, gobierno, entre otras) respecto de qué sustentante es apto para hacer uso de la lengua meta en diferentes

contextos: académicos, profesionales o con fines migratorios. Kane (2013, p. 46) retoma la postura de Bachman y Palmer (2010) y comenta que en la toma de decisiones deberá establecerse un balance entre las consecuencias positivas y negativas de las mismas, optando por aquellas en donde las positivas sobrepasen a las negativas.

2.5.3. EL MODELO DE VALIDACIÓN DE MISLEVY

Retomando la noción de argumentación de Toulmin (1958), Mislevy et al. (2003) desarrollaron el modelo de evaluación conocido como Diseño Centrado en la Evidencia (*Evidence-Centered Design*) para explicar el papel central que juega el razonamiento probatorio o a través de la evidencia. El modelo diseñado por los autores representó la inclusión por vez primera del argumento de validación para los exámenes de lengua y, al igual que el modelo de Bachman y Palmer (2010), resulta útil tanto para el diseño de exámenes como para la validación de los mismos.

Dicho modelo partió de la necesidad de poder establecer un argumento que permitiera relacionar lo que los estudiantes dicen o hacen en un examen de lenguas con lo que hacen en contextos más amplios. Por ello, el eje primordial del diseño de éste se encuentra sustentado en la noción de la Evaluación de Lengua Basada en Tareas (*Task-Based Language Assessment*) en las cuales se requiere la integración del conocimiento con las habilidades de la lengua (Mislevy et al., 2003, p. 5). Esta noción de tareas hace referencia a aquellas situaciones del “mundo real” en las cuales se hace uso de la lengua; mismas que fueron descritas por Bachman y Palmer (1996) como el Uso de la Lengua Meta (*Target Language Use*).

Según Mislevy et al. (2003), los pasos para el diseño de exámenes son equivalentes a los que se deben seguir en el proceso de validación de los mismos. Los autores incluyen cuatro pasos fundamentales, compuestos a su vez por diversos componentes cada uno de ellos: el análisis del dominio, el modelado del dominio, el marco conceptual de evaluación y la arquitectura de la prueba –la cual engloba: la aplicación y puntuación de la tarea– (Mislevy, 2006, 2009; Mislevy et al. 2003).

De manera muy sucinta se describen cada uno de los componentes del Diseño Centrado en la Evidencia. En primer lugar, es importante hacer un análisis del dominio; en otras palabras, de aquellas situaciones de la “vida real” para las cuales deseamos evaluar. En segundo término, los autores describen el modelado del dominio, el cual se conforma de las tareas y de todos los

instrumentos que se requerirán para evaluarlas; por ejemplo, las rúbricas. En tercera posición figura el marco conceptual de evaluación, en donde se encuentran todos aquellos procesos necesarios para determinar la confiabilidad de una prueba; llámense especificaciones, requerimientos operativos, modelos estadísticos, detalles de rúbricas, entre otros. Generalmente, aunque no siempre, dentro de este marco se encuentran todos los análisis estadísticos que se requieren y que son indispensables para determinar si los instrumentos que han sido diseñados, así como los mecanismos para medirlos, cumplen con la función para la cual fueron creados. Finalmente, la cuarta sección se encuentra representada por la arquitectura del examen, misma que incluye la evaluación operativa en la cual se considera el resultado de los exámenes y las partes involucradas: examinadores, administradores, examinandos, así como aquellos que toman las decisiones.

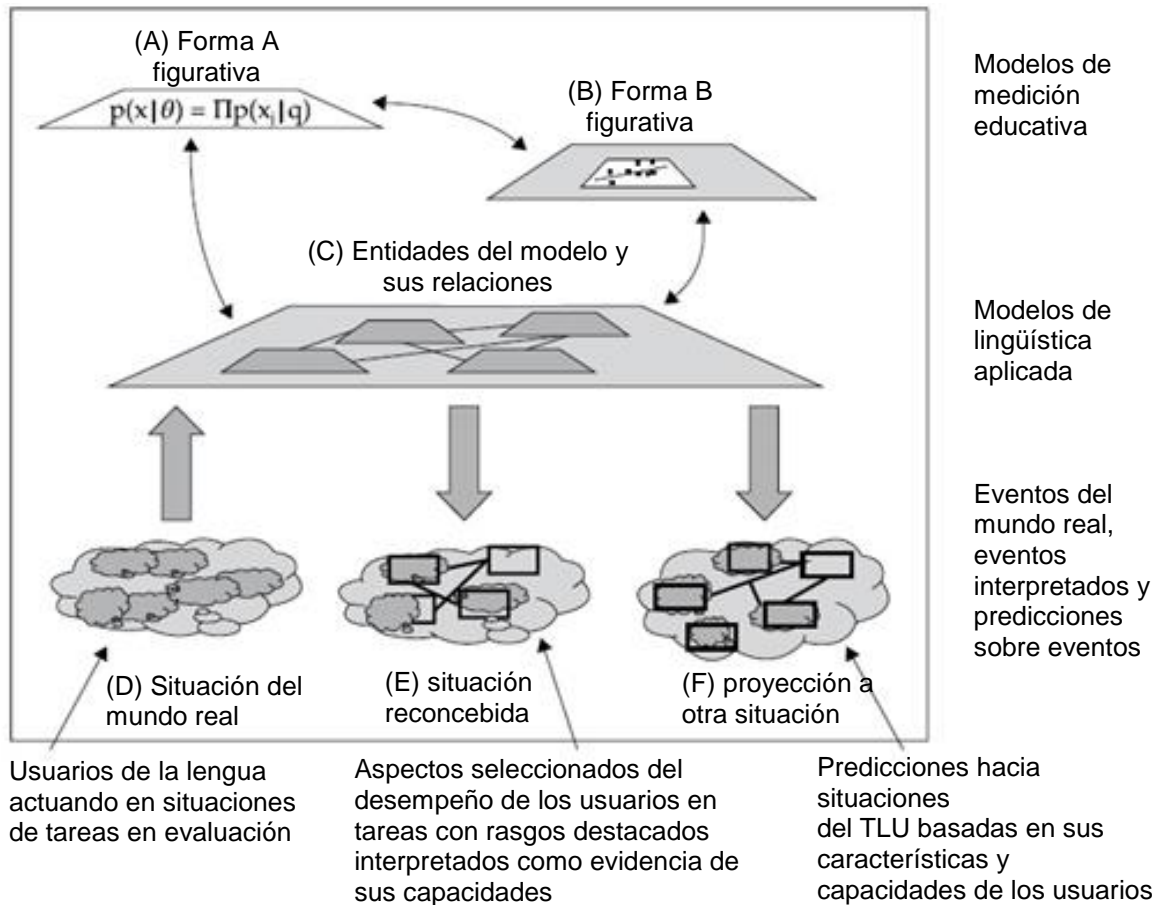
El modelo de Mislevy es bastante complejo y a través de los años ha sido reelaborado por el autor (2006⁴, 2009, 2013). No obstante, los puntos descritos anteriormente constituyen su base fundamental. En la actualidad, su modelo es conocido como Modelo Basado en el Razonamiento (*Model-Based Reasoning*) con sustento en una perspectiva sociocognitiva (Mislevy, 2013). Para Mislevy, un elemento esencial de la validez de una prueba resulta si al emplear un modelo, subyace una base sólida para la organización de las observaciones y para la guía de las acciones que deberán considerarse en las situaciones para las cuales el examen fue destinado.

La Figura 2.4 muestra el Modelo Basado en el Razonamiento de Mislevy. Como se puede apreciar, el autor concibe su modelo en tres planos. En la parte inferior se muestran diversas manifestaciones del mundo real (D, E y F). De acuerdo con Mislevy (2006, 2009, 2013), el sustento teórico y empírico son imprescindibles para poder diseñar las tareas mediante las cuales se evaluará a un examinando; por ello, la importancia del constructo dentro de la línea de validez de argumentación. El autor retoma el marco de Bachman y Palmer (1996) que incluye aspectos socioculturales de la lengua en uso para diseñar un examen. Ahora bien, para acceder del plano inferior al intermedio, según lo indica la flecha de la Figura 2.4, se deben considerar previamente los fundamentos teóricos y empíricos de los procedimientos necesarios para evaluar las tareas.

⁴ Mislevy (2006) reconoce la postura de Bachman (2005) respecto de la importancia del uso y consecuencias de los exámenes, incorporando el argumento de uso dentro de su discurso como parte del proceso de evaluación. A partir de esta postura, diseña una estructura para la evaluación de argumentos que consta de dos etapas: el argumento de evaluación y el argumento de uso.

Figura 2.4: Modelo basado en el razonamiento para la evaluación de lenguas (Mislevy, 2013, p. 2)

(Traducido por el investigador)



En el plano intermedio, denominado por Mislevy como “plano semántico” (C), se muestran las entidades, relaciones y propiedades del modelo; es decir, la serie de argumentos que tendrían que plantearse a manera de filtro para hacer inferencias y tomar decisiones respecto de las habilidades lingüísticas de un estudiante. Se trata entonces de un modelo de lingüística aplicada basado en la teoría y experiencia –el diseño de la prueba– que le den sustento científico a esta capa del modelo. Según Mislevy (2013), este plano intermedio, conformado por modelos lingüísticos, funge como intermediario entre el complejo “mundo real” (plano inferior) y el abstracto y rígido “mundo estadístico” que se emplea para validar exámenes (plano superior).

En el plano superior (A y B) figuran los sistemas de medición educativos que darán soporte al razonamiento y argumento previamente definido. Mislevy lo esquematiza a manera de fórmulas y ecuaciones que representan el proceso estadístico que debe efectuarse para poder

validar un examen y así, tomar decisiones. El analista, a través del modelo de razonamiento (E), conecta nuevamente el sistema científico de validez de la prueba con el “mundo real”, lo cual implicaría la toma de decisiones. Como se puede apreciar en el modelo, se parte de un examen concebido dentro del mundo real y el objetivo final es justamente la inferencia de lo que un sustentante será capaz de realizar en un contexto determinado para el cual está siendo examinado (F).

2.6. LA DISYUNTIVA ENTRE EVIDENCIA Y VERDAD

Los tres enfoques descritos anteriormente manifiestan que los argumentos deben encontrarse respaldados en todo momento mediante evidencia que le confiera credibilidad al argumento. No obstante, Borsboom y Markus (2013) sostienen que la verdad y la evidencia son diferentes, lo cual implica que aunque los exámenes cuenten con un respaldo plausible como evidencia, las interpretaciones que se generen de éstos podrían ser falsas.

Borsboom, Mellenbergh y Van Heerden (2004) mencionan que solamente los atributos que existen y que conocemos pueden ser medidos. Los autores agregan que si la variación de un atributo no causa variaciones en los resultados de los exámenes, entonces no estamos midiendo nada o estamos midiendo otra cosa diferente a lo que deseamos medir. De tal suerte que una de las cualidades de la validez se encuentra en la casualidad. Lo anterior, añaden los autores, pone en cuestionamiento el concepto de validez en cuanto a su enfoque epistemológico, su significado y su correlación. Por ello, Borsboom et al. (2004) puntualizan que el énfasis debería subyacer a una base ontológica, de significado y de causalidad (p. 1061). Entonces, según estos autores deberíamos partir de una simple premisa: un examen es válido si éste mide lo que pregona que mide. En este tenor, los autores cuestionan la razón por la cual se hace énfasis en las interpretaciones de los resultados de un examen (como es el caso de los enfoques basados en la argumentación previamente descritos) y no en las propiedades de éste, poniendo así en tela de juicio la validez propuesta por Cronbach y Meehl (1955) y también de Messick (1989). Más aún, critican que gran parte del énfasis en los exámenes se ha puesto en el análisis de éstos cuando en realidad el mayor énfasis debería recaer en la fase de construcción. Por ende, la fuente principal para comprender como funciona un examen habría de ser sustantiva y no metodológica (p. 1068).

Borsboom y Markus (2013) aseveran que la verdad y la justificación deben ir de la mano; sin embargo, la crítica que hacen al enfoque basado en la argumentación para validar exámenes

es que la argumentación puede llevar una conclusión falsa si las premisas sobre las cuales se encuentra sustentada son erróneas o falsas también. Por ende, contar con un argumento no es suficiente, sino que además sus premisas deben ser verdaderas. Aunado a lo anterior, Borsboom y Markus (2013) agregan que el mayor riesgo que se corre es recolectar información que dé sustento a un argumento para conferirle validez, aunque sus premisas sean falsas. De manera muy puntual, arguyen que uno puede generar un argumento en el cual el respaldo siempre garantizará la solidez de éste, puesto que fue diseñado con tal fin. Entonces, los autores proponen que la construcción de un argumento no sea el fin en sí mismo, sino el medio para entender qué tan bien funciona un examen (Borsboom y Markus, 2013, p. 111). Este tipo de argumento requiere entonces de la incorporación de la verdad y de la justificación como elementos distintos. En suma, lo que los autores exponen es que el enfoque de la argumentación vincula la interpretación con la justificación en vez de hacerlo de forma paralela con la justificación y la verdad, lo cual posiciona a la validez como independiente de la verdad y que, por lo tanto, es susceptible de ser falsa.

2.7. DISCUSIÓN SOBRE LOS ENFOQUES DE ARGUMENTACIÓN PARA VALIDAR EXÁMENES

Los tres enfoques descritos anteriormente han aportado al proceso de validación de exámenes y se han retroalimentado a lo largo de más de una década. Mislevy et al. (2003) fueron los primeros en incorporar el concepto de argumentación de Toulmin en el proceso de validación de exámenes de lengua. En 2006, Mislevy incorporó la noción de argumento de uso propuesto por Bachman (2005); sin embargo, la manera mediante la cual pudiera emplearse en el proceso de validación no es clara. De hecho, su modelo es complejo y hace gran énfasis en los modelos matemáticos que deberán utilizarse para lograr la confiabilidad de la prueba, lo cual dificulta su aplicación y deja del lado inferencias importantes que deben ser tomadas en cuenta al momento de validar un examen. Otra de las desventajas que el mismo Mislevy reconoce (2006), respecto de los modelos, es que éstos no son perfectos y resulta difícil adaptarlos, ya que no siempre es claro cuáles podrían ser los criterios para flexibilizarlos al momento de emplearlos.

La mayor aportación del modelo de Bachman (2005) y Bachman y Palmer (2010) se debe al denotado énfasis en las consecuencias de los exámenes. De hecho, el argumento de validación busca como fin último el uso que se hará de éstos. La crítica de Bachman y Palmer (2010) hacia los modelos de Kane y Mislevy reside fundamentalmente en que estos autores no incorporan el concepto de uso (Kane, 2006) o lo hacen de manera tangencial (Mislevy, 2006).

Por ello, Kane (2013) incorpora dentro de su discurso el argumento de interpretación y uso de los exámenes. No obstante, a diferencia del peso mayoritario que le confiere Bachman (2005) y Bachman y Palmer (2010), Kane afirma que ambos argumentos (de interpretación y uso) son igual de importantes. Pese a lo anterior, la línea discursiva de Bachman (2005) y Bachman y Palmer (2010) hace referencia profusa a las consecuencias y al uso de los exámenes, tomando en cuenta las nociones de equidad y de ética, pero dejando del lado la importancia de las otras inferencias que son fundamentales para crear un argumento sólido en cada uno de los estadios que se requiere para llegar de los datos a la aserción.

El enfoque de Kane, por otra parte, es el que mayor desarrollo, profusión y aplicaciones ha tenido desde la década los 90 cuando incorporó por vez primera el argumento por el cual Cronbach (1988) y Messick (1989) abogaban con el fin de concebir la validez de forma unitaria. Su enfoque incorporó desde sus inicios el modelo argumentativo de Toulmin para la validación de exámenes. Sin embargo, debido a que el modelo de Toulmin no fue diseñado específicamente para la validación de exámenes de lenguas, no fue sino hasta comienzos del siglo XXI que fue adoptado con tal fin. Como bien señala Kane (2013), una de las mayores ventajas de su enfoque se debe a que es flexible y moldeable a la situación para la cual se pretenda validar un examen. Las inferencias que el autor propone no resultan a discreción del experto en evaluación, sino más bien de acuerdo con las necesidades para las cuales se pretende evaluar. Por ello, el autor describe detalladamente en qué casos será conveniente incluir cada una de las inferencias que él propone, así como la evidencia que sería necesaria para fundamentarlas.

Otra de las ventajas del enfoque de Kane se debe precisamente a que al igual que Toulmin, su marco se encuentra sujeto a ser refutado mediante la contraevidencia. Chapelle, Enright y Jamieson (2008) advierten que sin lugar a dudas esto representa una de las grandes ventajas del enfoque, pues precisamente el objetivo de la validez no es aceptar y aprobar un modelo de examen, sino encontrar fallas y huecos que permitan su perfeccionamiento, desarrollo y evolución. Por esta razón, el enfoque de Kane se ha adaptado a las condiciones que imperan en materia de evaluación y también de validez de exámenes. Finalmente, el reconocimiento de Kane (2013) respecto del uso y las consecuencias de los exámenes, además de la interpretación de sus resultados, le confieren a su enfoque la solidez que se requiere cuando se trata de exámenes que conllevan implicaciones importantes en la vida de los examinandos; en este caso, de los estudiantes no nativohablantes que desean ingresar a la universidad (Fulcher y Davidson, 2007;

McNamara y Roever, 2006), asimismo, implica un reconocimiento de la importancia que actualmente juega la ética y la equidad en la toma de decisiones.

En cuanto a la crítica de Borsboom y Markus (2013) hacia la falta de verdad en la construcción de un argumento de validación de exámenes, para que en realidad ésta pudiera ser incluida deberíamos ser capaces de evidenciar claramente cuáles son las competencias que posee un sustentante. Sin embargo, no existe aún ninguna teoría que dé cuenta de cuáles son las competencias comunicativas en lenguas extranjeras, y como afirma Bachman (2006), éstas no son visibles. Pese a esto, vale la pena rescatar la visión de Borsboom y Markus (2013) sobre verdad para tener claridad de que más que la búsqueda de ésta, es importante recalcar que el proceso de validación de exámenes no se sustenta en ésta. Bachman y Palmer (2010) y Kane (2013) aseveran que debemos poner especial énfasis en que los resultados sean útiles para las partes interesadas y que las consecuencias benéficas sobrepasen aquellas perjudiciales.

De tal suerte que, pese a que no podamos conocer a detalle las competencias comunicativas de quienes sustentan exámenes de lengua, el enfoque de la argumentación en la validación de exámenes representa un gran apoyo para quienes se dedican a validar exámenes de lengua; no obstante, el verdadero reto se presenta al momento de generar un argumento plausible para validarlos. Como se mencionó anteriormente, aun cuando el modelo teórico de Messick (1989) constituyó un sólido cimiento en la concepción unitaria de validez, la problemática de aplicación y de la aún latente visión fragmentaria para validar exámenes que persistió hasta principios del siglo XXI (AERA, APA Y NCME, 1985, 1989; Alderson, Clapham y Wall, 1995; Bachman, 1990; Bachman y Palmer, 1996; Cumming y Berwick, 1996; McNamara 1996; Weir, 2005) dificultó en gran medida abordar la compleja tarea de validar exámenes de lengua. Por ello, el reto que actualmente subyace al proceso de validar exámenes consiste en la elaboración de un argumento que se encuentre respaldado mediante evidencia, que sea sujeto a refutación y que sea perfectible en sus secciones más endeables.

2.8. RESUMEN

A partir de la segunda mitad del siglo XX, la evaluación de lenguas adquirió un carácter científico y la validez de constructo constituyó el eje rector para garantizar la fiabilidad de los exámenes. No obstante, como se pudo apreciar, debido a las dificultades que se presentaron a finales del siglo XX relacionadas con la materialización de los modelos de validez en la

evaluación, incorporada de manera fragmentada y desarticulada, se ha adoptado desde inicios del siglo XXI la argumentación en la validez vista de manera integradora a través de la lógica argumentativa de Toulmin. De ahí que, el establecimiento de una asección respecto del desempeño de un candidato requiera de una serie de inferencias y su respectiva evidencia con el fin de garantizar la confiabilidad y validez de un examen. Los tres enfoques descritos demuestran el esfuerzo que se ha llevado a cabo durante la última década con miras a garantizar una evaluación justa y equitativa de los exámenes, particularmente en aquellos de alto impacto, puesto que conllevan importantes implicaciones en la vida de los examinandos. Dentro de estos enfoques, destaca el de Kane, mismo que gracias a su flexibilidad ha ido evolucionando a lo largo de casi dos décadas desde su concepción inicial, y se ha ido adaptando de forma natural de acuerdo con las nuevas exigencias e interrogantes que surgen en materia de evaluación. Sin lugar a dudas, la concepción de la evaluación vista desde la óptica de la lógica argumentativa resulta una herramienta útil y novedosa para el diseño y validación de exámenes de lenguas extranjeras.

En el siguiente capítulo, se describirá el argumento que se diseñó para validar la PE del EXELEAA y se esbozarán los estudios que se llevaron a cabo en la presente investigación para respaldar dicho argumento.

CAPÍTULO 3

EL ARGUMENTO DE VALIDACIÓN DE LA PE DEL EXELEAA

3.1 INTRODUCCIÓN

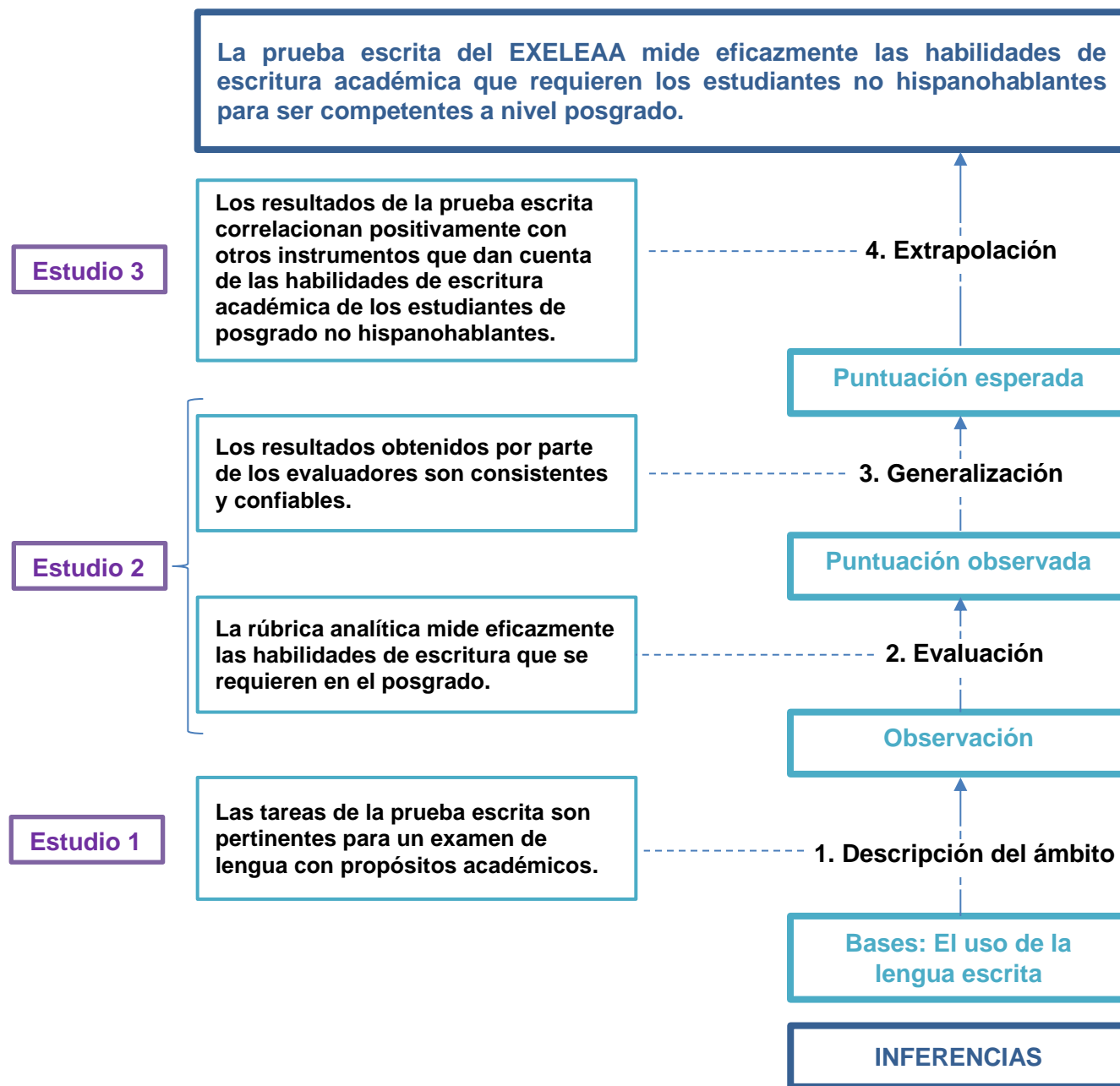
En este capítulo se presenta el argumento de validación de la PE del EXELEAA. Este argumento se diseñó siguiendo el enfoque de argumentación de Kane (2002b, 2006, 2013) y Kane et al. (1999) para validar exámenes. Posteriormente, se hará una descripción detallada de las inferencias abordadas en la presente investigación (i.e., descripción del ámbito, evaluación, generalización y extrapolación) y de los estudios que se diseñaron para respaldar cada una de las aserciones planteadas en cada una de las distintas inferencias. Finalmente, se presenta el esquema de cada uno de los tres estudios que se llevaron a cabo en la presente investigación. Cada estudio es independiente entre sí, aunque tangencialmente se enlazan con el fin de dar coherencia y cohesión al argumento de validación de la PE del EXELEAA.

3.2. DISEÑO DEL ARGUMENTO DE VALIDACIÓN DE LA PE DEL EXELEAA

Siguiendo como modelo el esquema de Chapelle (2012) y Chapelle, Enright y Jamieson (2008, 2010) para diseñar el argumento de validación del examen del TOEFL *Internet-Based Test* (iBT), y basándonos en el enfoque del argumento de validación de Kane (2002b, 2006, 2013), se diseñó el argumento de validación para la PE del EXELEAA. Para poder afirmar o refutar una aserción, Kane (2013) propone cuatro inferencias que deberán sustentarse mediante evidencia. Estas cuatro inferencias son: evaluación, generalización, extrapolación y utilización. Las tres primeras son de tipo semántico y la cuarta de tipo político. Sin embargo, debido a la extensión de esta investigación, se abordaron exclusivamente las de tipo semántico. Es decir, en el presente estudio no se contempla el uso que se le da al examen; no obstante, sería muy importante abordar dicha inferencia política en futuras investigaciones, pues el fin último de un examen reside en el uso que se le da. Kane sugiere que cuando se trata de exámenes de alto impacto, en donde la toma de decisiones podría afectar decisivamente la vida de los examinandos, se incluya la inferencia de la descripción del ámbito, misma que permitirá identificar qué tan pertinentes resultan las tareas seleccionadas para generalizar del ámbito para el cual se pretende evaluar. En la Figura 3.1 se presenta el argumento de validación de la PE del EXELEAA, mismo que consta de las cuatro inferencias previamente enunciadas.

Figura 3.1: Argumento de validación de la PE del EXELEAA

(Elaborado por el investigador)



En cada una de las cuatro inferencias se establecieron aseercciones que requirieron de evidencia para respaldar el argumento de validación. Por esta razón, la presente investigación se encuentra segmentada en tres estudios que se diseñaron para dar cuenta de cada una de las inferencias establecidas. El Estudio 1 aborda la inferencia de la descripción del ámbito y la

pertinencia de las tareas incluidas en la PE del EXELEAA. El Estudio 2 se centró en dos inferencias que se encuentran íntimamente vinculadas y que se relacionan con el proceso de corrección de la PE: la evaluación y la generalización. Finalmente, en el Estudio 3 se respaldó la inferencia de la extrapolación. Los Estudios 1 y 3 pretenden respaldar la validez externa del EXELEAA, mientras que el Estudio 2 se centra en su validez interna. Como consecuencia, cada uno de los estudios se conformó por distintas fases que pudieran dar cuenta de las aserciones establecidas en cada una de las inferencias. De ahí que, la investigación siguiera un método mixto. A continuación, se describirá el diseño de cada uno de los estudios contemplados en la presente investigación.

3.3. DISEÑO DE LA INVESTIGACIÓN

Creswell y Plano Clark (2011) mencionan que los métodos mixtos combinan elementos cuantitativos y cualitativos para poder tener un entendimiento más profundo de un fenómeno. Asimismo, las debilidades de un método se ven fortalecidas por el otro método y viceversa. En este tipo de estudios, se recolectan y analizan datos cuantitativos y cualitativos con el fin de poder establecer ciertas inferencias. Los autores comentan que en este tipo de estudios el investigador:

- Recolecta y analiza rigurosamente la información cuantitativa y cualitativa de acuerdo con las preguntas de investigación.
- Mezcla o integra los dos métodos de forma concurrente al combinarlos o yuxtaponerlos.
- Da prioridad a uno u otro, o ambos, dependiendo del énfasis de la investigación.
- Usa estos procedimientos en un estudio simple o en múltiples fases de un estudio.
- Enmarca los procedimientos bajo ciertos esquemas filosóficos o teóricos.
- Combina los procedimientos en diseños específicos dependiendo del esquema de la investigación (p. 5).

En la presente investigación, se optó por un estudio mixto de tipo convergente (Creswell y Plano Clark, 2011, p. 180); es decir, la recolección de los datos cuantitativos y cualitativos se hizo de forma concurrente, a excepción del Estudio 1, pero se analizaron de forma separada y posteriormente se combinaron los resultados cuantitativos y cualitativos. Al igual que en la propuesta de Creswell y Plano Clark, el estudio cualitativo que aquí se presenta tuvo como objetivo darle seguimiento a un pequeño grupo de individuos que participó en los estudios cuantitativos con el fin de poder extrapolar y explicar los resultados obtenidos en la PE. En suma, la combinación de estudios cuantitativos y cualitativos proporciona un entendimiento más

completo de un problema de investigación al combinar un enfoque de pensamiento inductivo con uno deductivo.

Dependiendo de la inferencia que se deseó respaldar, se llevaron a cabo diversos estudios cuantitativos y cualitativos en tres estudios: la descripción del ámbito y la pertinencia de las tareas de la PE del EXELEAA, la calidad proceso de corrección de la prueba escrita y la extrapolación de los resultados de la PE a las habilidades de escritura dentro del ámbito académico. En el Cuadro 3.1 se muestra el esquema general de la investigación, así como de los 3 estudios de los cuales se compuso.

Cuadro 3.1: Esquema general de la investigación

Investigación mixta de tipo convergente	
<p>Estudio 1</p> <p>Inferencia 1: descripción del ámbito</p>	<p>Tipo de estudio:</p> <ul style="list-style-type: none"> • exploratorio de tipo descriptivo • mixto: cuantitativo y cualitativo • instrumento 1: cuestionario abierto • instrumento 2: encuesta en escala tipo Likert • administración en línea • recuento de respuestas en una tabla en EXCEL, instrumento 1 • análisis estadístico en SPSS, instrumento 2 <p>Participantes:</p> <ul style="list-style-type: none"> • docentes del posgrado de la UNAM <p>Objetivos:</p> <ul style="list-style-type: none"> • conocer las prácticas de evaluación de los docentes a nivel posgrado. • identificar las prácticas de escritura a nivel posgrado, así como su importancia con respecto a otras formas de evaluación. • investigar la importancia de las habilidades de escritura académica a nivel posgrado.
<p>Estudio 2</p> <p>Inferencias 2 y 3: evaluación y generalización</p>	<p>Parte 1</p> <p>Tipo de estudio:</p> <ul style="list-style-type: none"> • cuantitativo • instrumento: rúbrica analítica para evaluar los textos de la PE del EXELEAA • análisis estadístico en Many-Facet Rasch Measurement (MFRM) <p>Participantes:</p> <ul style="list-style-type: none"> • estudiantes universitarios no hispanohablantes • correctores de pruebas escritas <p>Objetivo:</p> <ul style="list-style-type: none"> • conocer la calidad del proceso de evaluación de la PE del EXELEAA; a saber: la dificultad de las tareas, la severidad de los evaluadores y el funcionamiento de la rúbrica analítica.

<p style="text-align: center;">Estudio 2</p> <p>Inferencias 2 y 3: evaluación y generalización</p>	<p>Parte 2 Tipo de estudio:</p> <ul style="list-style-type: none"> • cualitativo • instrumento: cuestionario abierto • recuento de respuestas en una tabla de EXCEL <p>Participantes:</p> <ul style="list-style-type: none"> • correctores de pruebas escritas <p>Objetivo:</p> <ul style="list-style-type: none"> • conocer las percepciones y la opinión de los correctores respecto de la rúbrica analítica empleada para evaluar los ensayos de la PE del EXELEAA, así como de proceso de capacitación. <p>Parte 3 Tipo de estudio:</p> <ul style="list-style-type: none"> • cuantitativo • instrumentos: rúbrica analítica y listado de las habilidades de escritura identificadas por docentes como importantes a nivel posgrado <p>Participantes:</p> <ul style="list-style-type: none"> • docentes con experiencia en escritura académica y en el diseño y uso de rúbricas analíticas empleadas con no hispanohablantes <p>Objetivo:</p> <ul style="list-style-type: none"> • identificar las habilidades de escritura académica que son evaluadas por la rúbrica analítica empleada en la PE del EXELEAA.
<p style="text-align: center;">Estudio 3</p> <p>Inferencia 4: extrapolación</p>	<p>Parte 1 Tipo de estudio:</p> <ul style="list-style-type: none"> • cuantitativo • instrumento 1: encuesta en escala tipo Likert • administración en línea • análisis estadístico correlacional en SPSS • instrumento 2: rúbrica analítica para evaluar entrevistas <p>Participantes:</p> <ul style="list-style-type: none"> • estudiantes de posgrado no hispanohablantes • evaluadores de entrevistas <p>Objetivos:</p> <ul style="list-style-type: none"> • correlacionar los resultados de la encuesta sobre habilidades de escritura académica de los estudiantes de posgrado no hispanohablantes, el resultado obtenido al evaluar la entrevista sobre habilidades de escritura académicas y el resultado en la PE del EXELEAA. <p>Parte 2 Tipo de estudio:</p> <ul style="list-style-type: none"> • cualitativo • instrumento: guía semiestructurada sobre el proceso de escritura académica y variables asociadas a éste. • entrevista presencial • análisis cualitativo en NVivo

<p>Estudio 3</p> <p>Inferencia 4: Extrapolación</p>	<p>Participantes:</p> <ul style="list-style-type: none"> • estudiantes de posgrado no hispanohablantes • codificadores de datos cualitativos <p>Objetivo:</p> <ul style="list-style-type: none"> • investigar el proceso de escritura académica de los estudiantes de posgrado no hispanohablantes y las variables asociadas a dicho proceso.
--	--

El Estudio 1 tuvo como objetivo determinar la pertinencia de las tareas escritas contempladas en la PE del EXELEAA. Para ello, lo primero fue conocer el ámbito para el cual deseamos evaluar, así como las tareas de escritura que comúnmente son empleadas para evaluar las habilidades de escritura de los estudiantes de posgrado. Sin embargo, debido a que el campo de la escritura académica es vasto, se condujo un estudio exploratorio para conocer solamente dos aspectos del ámbito: la importancia de la escritura a nivel posgrado y la importancia de las habilidades de escritura que requieren los estudiantes de posgrado para ser académicamente competentes. Para esta primera fase del estudio, se elaboró un cuestionario abierto con el fin de conocer cuáles son las prácticas de evaluación de los docentes del posgrado de la UNAM. Asimismo, estos mismos docentes participaron en una encuesta sobre cuáles son las habilidades de escritura más importantes para los estudiantes del posgrado. La información recabada permitió que se determinara la pertinencia de las tareas de la PE del EXELEAA para el ámbito académico.

El Estudio 2, compuesto por estudios cuantitativos, se diseñó para dar cuenta de la calidad del proceso de corrección. Las tareas que se incluyen en un examen determinan en gran medida la manera mediante la cual interactúa el candidato con el texto. Este proceso de interacción deviene en la dificultad de la tarea y, consecuentemente, en el desempeño del sustentante. En el proceso de corrección de los textos, las escalas de evaluación y los evaluadores juegan un papel indisoluble. De ahí que, la asignación de un puntaje en las pruebas escritas dependa de un complejo proceso de interacción generado entre el evaluador, la rúbrica y el texto. Por ello, es indispensable contar con herramientas estadísticas que permitan evidenciar el funcionamiento de la rúbrica y también determinar la consistencia y confiabilidad de los resultados otorgados por los evaluadores. Esto garantiza que los resultados obtenidos por un examinado sean justos y que correspondan eficazmente a su desempeño en lengua meta. Para poder llevar a cabo esta tarea, fue necesario explorar la calidad de la rúbrica en tres diferentes aspectos: el funcionamiento de la escala, la percepción de los evaluadores respecto de la rúbrica y la factibilidad de la rúbrica de medir las habilidades de escritura académica requeridas a nivel

posgrado. Debido a que tanto las tareas de la PE como las rúbricas para evaluarlas eran de diseño reciente, esta parte del estudio se dividió en dos etapas: pilotaje y corrección con la rúbrica modificada.

Finalmente, el Estudio 3 abordó la extrapolación de los resultados de la PE al “mundo real”. La evaluación no constituye un fin, sino el medio a través del cual se puedan tomar decisiones respecto del desempeño de los examinados. En el caso de los ECLAA, la decisión consiste en admitir o no a los estudiantes prospectos con una lengua materna diferente a la del medio de instrucción. De hecho, se infiere que un examinado que demuestra ciertas habilidades; por ejemplo, de escritura, será capaz de escribir eficientemente dentro de una amplia gama de tareas y trabajos académicos que le serán requeridos durante sus estudios universitarios. No obstante, esta generalización también debe encontrarse sustentada mediante estudios formales que evidencien que, efectivamente, entre mayor sea el resultado obtenido por un examinado en la sección escrita de un examen de segundas lenguas, mayores serán sus habilidades de escritura y su capacidad para escribir exitosamente en el ámbito académico. Contrariamente, también deberá confirmarse lo opuesto; es decir, que quienes obtienen un puntaje bajo en la PE, también presentan deficiencias y tropiezos para escribir adecuadamente en el contexto universitario. Así que, validar la inferencia de la competencia comunicativa (en este caso de escritura) de un examinado en otros contextos que no se circunscriben a los del examen resulta crucial dentro del proceso de validación. Con el fin de poder llevar a cabo el proceso de comparación, se diseñó una entrevista semiestructurada para poder evidenciar las habilidades de escritura de los estudiantes en el “mundo real”. Mediante una rúbrica analítica, se evaluó la habilidad de escritura de los estudiantes. Asimismo, se les administró una encuesta en escala tipo Likert para que autoevaluaran sus habilidades de escritura académica. Para sustentar la inferencia de extrapolación, los datos de estos dos instrumentos se correlacionaron con los resultados que obtuvieron los estudiantes en la PE del EXELEAA.

En los Capítulos 4, 5 y 6, se describirán los tres estudios que conformaron la presente investigación.

CAPÍTULO 4

ESTUDIO 1: LA DESCRIPCIÓN DEL ÁMBITO ACADÉMICO Y LAS TAREAS DE LA PE DEL EXELEAA

4.1. INTRODUCCIÓN

En los exámenes de certificación de lengua para ámbitos específicos (ECLAE) es importante conocer el contexto para el cual se desea evaluar. A diferencia de los exámenes de certificación de lengua para ámbitos generales (ECLAG), en los ECLAE sabemos cuál será el uso que se le dará al examen. Por ejemplo, los ECLAA más comunes son aquellos cuyo resultado se toma como criterio de ingreso para el ciclo universitario de grado o de posgrado en medios de instrucción distintos a la lengua materna del prospecto universitario. Por este motivo, es importante conocer cuál es el uso de la lengua en dicho ámbito, en aras de construir un examen que incluya tareas del uso de la lengua en el contexto para el cual se evalúa. Este tipo de trabajo no es sencillo, pues los exámenes representan una pequeña muestra –“universo de generalización”– del uso de la lengua en el ámbito para el cual se evalúa. Las tareas que se incluyen en un examen determinan en gran medida la manera mediante la cual interactúa el candidato con el texto que produce. Este proceso de interacción deviene en la dificultad de la tarea y, consecuentemente, en el desempeño del sustentante.

4.2. MARCO TEÓRICO

Dado que en la UNAM no se cuenta con un Marco Curricular Universitario en el que se definan las competencias comunicativas que deberán poseer los estudiantes, en este apartado se hará una descripción detallada de las habilidades de escritura y los géneros discursivos que frecuentemente se empelan dentro del ciclo universitario para evidenciar el conocimiento que han adquirido los estudiantes. Posteriormente, se revisarán los distintos tipos de tareas escritas que se solicitan en los ECLAA para que el examinado demuestre la habilidad con que cuenta para redactar textos que se encuentran relacionados con el ámbito académico.

4.2.1. LA DESCRIPCIÓN DEL ÁMBITO ACADÉMICO

Kane (2013) sugiere que en los exámenes de alto impacto se describa cuidadosamente el ámbito para el cual deseamos evaluar. Este tipo de conocimiento del uso de la lengua no es parte del examen, pero sí es relevante, pues representa las bases del constructo sobre las cuales se concebirá y diseñará el examen. El gran problema que emerge al momento de determinar cuáles son las habilidades lingüísticas que debería poseer un estudiante universitario radica en que frecuentemente éstas no se encuentran enunciadas en los planes curriculares de los programas universitarios, ni tampoco en los perfiles de ingreso o egreso. En el caso del Bachillerato mexicano, por ejemplo, se cuenta con el Marco Curricular para la Educación Media Superior, el cual pretende homogenizar y estandarizar las competencias requeridas por los estudiantes de este ciclo escolar. Sin embargo, en el caso del nivel superior, no existe un documento homólogo que enuncie tales competencias.

No obstante, debido a la alta movilidad estudiantil y laboral a nivel mundial, favorecida por la globalización, se ha generado la necesidad de encontrar un espacio de convergencia entre los sistemas educativos que permitan dar cuenta de las competencias que posee un estudiante universitario que desea estudiar o trabajar en un país extranjero. Tales demandas dieron origen a la construcción de un proyecto internacional llamado *Tuning*. Éste fue creado originalmente para que las universidades europeas “pusieran en sintonía” sus estructuras educativas universitarias con el fin de que los títulos universitarios pudieran ser reconocidos a lo largo de la Comunidad Europea. Su contraparte latinoamericana¹, de manera análoga, tiene como finalidad “buscar puntos de acuerdo, convergencia y entendimiento mutuo para facilitar la comprensión de las estructuras educativas” (Beneitone et al. 2007, p. 11). *Tuning* América Latina retoma y adapta las 30 competencias genéricas utilizadas por su contraparte europea. De manera muy sucinta, se enuncian aquellas relacionadas con la expresión escrita.

- 1) Capacidad de abstracción, análisis y síntesis.
- 6) Capacidad de comunicación oral y escrita.

¹ Consúltase Proyecto *Tuning* (2007). Informe Final del Proyecto *Tuning* América Latina: Reflexiones y perspectivas de la Educación Superior en América Latina
Recuperado del portal de Internet:

http://tuning.unideusto.org/tuningal/index.php?option=com_docman&Itemid=191&task=view_category&catid=22&order=dmdate_published&ascdesc=DESC

Última consulta el 26 de diciembre de 2016.

11) Habilidades para buscar, procesar y analizar la información procedente de diversas fuentes

12) Capacidad crítica y autocrítica.

(Beneitone et al., p. 44)

Como se puede observar, dentro de estas competencias, se destacan las comunicativas. En cuanto a las habilidades de escritura dentro del ámbito académico, destacan aquellas que se vinculan con la construcción del conocimiento, dado que se trata de una habilidad necesaria para evidenciar y evaluar el conocimiento de los estudiantes universitarios; por ejemplo, mediante un trabajo escrito perteneciente a un género discursivo en particular, una tesis de investigación o artículos científicos. De hecho, la producción de textos científicos y académicos se ha ampliado significativamente entre la comunidad estudiantil del posgrado. En este sentido, cabe agregar que cada vez es más frecuente, por lo menos en el contexto de la UNAM, que las facultades les exijan a sus alumnos de posgrado que publiquen artículos científicos como requisito de titulación. Dependiendo del posgrado, muchas veces se solicita que la publicación sea incluso en inglés.

Debido a que los ECLAA se encuentran íntimamente relacionados con el ámbito para el cuál se evalúa, resulta necesario conducir estudios que evidencien las competencias comunicativas que requieren los prospectos no hispanohablantes a cursar estudios de posgrado. Solamente de esa manera, se podrán diseñar exámenes que incluyan tareas auténticas y representativas del ámbito para el cual se desea evaluar. Sin embargo, no se han explorado de manera exhaustiva en el contexto hispanohablante cuáles son estas habilidades que requieren los estudiantes a nivel posgrado. Más aún, agregan Castelló, González e Iñesta (2010) que no se sabe cuáles son las dificultades por las que atraviesan los estudiantes de posgrado al momento de escribir una tesis, ni tampoco sobre las estrategias o recursos adicionales que emplean para solventarlas.

4.2.1.1. LAS HABILIDADES DE ESCRITURA ACADÉMICA

Debido a la amplia gama de recursos escritos que se emplean a nivel posgrado para la construcción y difusión del conocimiento, la escritura académica en este ciclo escolar se ha especializado y se ha convertido en una herramienta necesaria para que los estudiantes de posgrado sean académicamente exitosos. No obstante, los estudios que se han conducido para investigar la importancia de las habilidades de escritura en el ámbito académico son escasos. Destacan, sin embargo, los estudios a gran escala que Rosenfeld, Leung y Oltman (2001) y

Rosenfeld, Courtney y Fowles (2004) llevaron a cabo dentro del contexto anglófono y recientemente, en el contexto hispanohablante, destaca el estudio de Castelló, Mateos, Castells, Iñesta, Cuevas y Solé (2012).

En el primer estudio, los investigadores identificaron las habilidades de comprensión auditiva y lectora, así como de expresión escrita y oral que requieren los estudiantes para ser académicamente exitosos. Si bien es cierto que el objetivo primordial de dicha investigación fue recabar información para el diseño de las tareas del nuevo examen de dominio de inglés, TOEFL iBT, la información obtenida resultó importante para identificar las competencias lingüísticas que requieren los estudiantes de grado y de posgrado.

En un estudio más fino, Rosenfeld et al. (2004) se dieron a la tarea de identificar las habilidades de escritura que los docentes consideran importantes para que los estudiantes de grado y de posgrado sean académicamente competentes. Los investigadores organizaron varias sesiones con especialistas en el área y después de varios procesos de depuración, seleccionaron 39 descriptores de dichas habilidades. El Cuadro 4.1 muestra los resultados de las habilidades de escritura que Rosenfeld et al. (2004) identificaron como más importantes dentro del contexto universitario. Como se puede apreciar, estas habilidades se relacionan con la construcción del conocimiento (i.e., dar crédito apropiado a las fuentes, analizar y sintetizar información de múltiples fuentes, integrar el material citado en un texto y abstraer o resumir información), la organización de ideas de forma coherente, clara y precisa, el uso correcto de la gramática, y la edición y revisión de textos.

En el contexto hispanohablante, aunque aún con estudios incipientes en este terreno, destaca el estudio de Castelló et al. (2012) para identificar las habilidades de escritura que los docentes de cuatro universidades españolas consideraron importantes para el ámbito académico. Según estos autores, los criterios que resultan imprescindibles para evaluar la calidad de los textos académicos son: claridad, organización y precisión, terminología precisa, claridad de puntos, uso apropiado del lenguaje y de la revisión de la literatura. En el caso de los estudios doctorales, Castelló e Iñesta (2012) destacan que la revisión del texto resulta una habilidad indispensable para ser académicamente competente. En suma, agregan Castelló et al. (2012), los maestros, expertos en escritura y estudiantes avanzados consideran que los textos académicos deben ser claros y precisos en su estructura y lógica textual para ser asequibles al

lector. Como se puede apreciar, los hallazgos en cuanto a las habilidades de escritura en el contexto hispanohablante encuentran puntos de convergencia con su contraparte anglófona.

Cuadro 4.1: Habilidades de escritura más importantes a nivel maestría
(Rosenfeld et al., 2004, p. 14 y 15)
(Traducido por el investigador)

Habilidades de escritura	Puntuación media
Dar crédito adecuado a las fuentes.	4.5
Organizar las ideas y la información coherentemente.	4.4
Utilizar correctamente la gramática del inglés escrito con el fin de evitar que el lector se distraiga o que se cometan errores que alteren el significado.	4.4
Evitar errores ortográficos.	4.3
Abstraer o resumir información esencial.	4.2
Analizar y sintetizar información de múltiples fuentes (incluye comparación y contraste).	4.2
Integrar el material citado y referenciado debidamente dentro del texto.	4.2
Desarrollar un debate bien enfocado y debidamente sustentado mediante ejemplos relevantes.	4.2
Escribir con claridad mediante transiciones suaves de un pensamiento a otro.	4.2
Escribir de forma precisa y concisa, evitando frases vagas o vacías.	4.2
Revisar y editar el texto para mejorar su calidad, coherencia y precisión.	4.2

4.2.1.2. LOS GÉNEROS DISCURSIVOS A NIVEL POSGRADO

Sin embargo, las habilidades de escritura que requiere un estudiante de posgrado son dependientes de la comunidad discursiva a la cual pertenecen. Por esta razón, uno de los escollos que se han suscitado al momento de conducir estudios relacionados con la escritura académica radica en la amplia gama de géneros discursivos y en las diferencias de lo que se entiende por un mismo género discursivo no sólo entre diversas comunidades científicas, sino entre los mismos docentes pertenecientes a una misma disciplina académica. Entonces, la dificultad de caracterizar los diversos géneros discursivos ha dificultado aún más los estudios que se conducen sobre las habilidades de escritura que se requieren en el ámbito académico.

Pese a que la sistematización de las tipologías textuales en el ámbito académico resulta una tarea intrincada, en el contexto anglófono hay estudios que dan cuenta de éstas (Canseco y Byrd, 1989; Cooper y Bikowski, 2007; Gardner y Nesi, 2012; Hale, Taylor, Bridgeman, Carson, Kroll y Kantor, 1996; y Horowitz, 1986). Sin embargo, en su contraparte hispanohablante, estas investigaciones son escasas (Castelló et al., 2012). En el caso anglófono, las investigaciones han seguido dos rumbos diferentes. El Cuadro 4.2 muestra un resumen de las tipologías

textuales identificadas en el contexto anglófono. Algunas de ellas, mediante un análisis textual, han recolectado y analizado textos de los estudiantes con miras a clasificarlos de acuerdo a las convenciones que se emplean. Otros estudios han recolectado información de diversos programas de cursos universitarios con el fin de identificar aquellas tipologías textuales que normalmente solicitan los docentes para evaluar a los estudiantes. Sin importar cuál haya sido el acercamiento, estos estudios muestran la importancia de los géneros textuales a los cuales se enfrentan los estudiantes universitarios de grado y de posgrado.

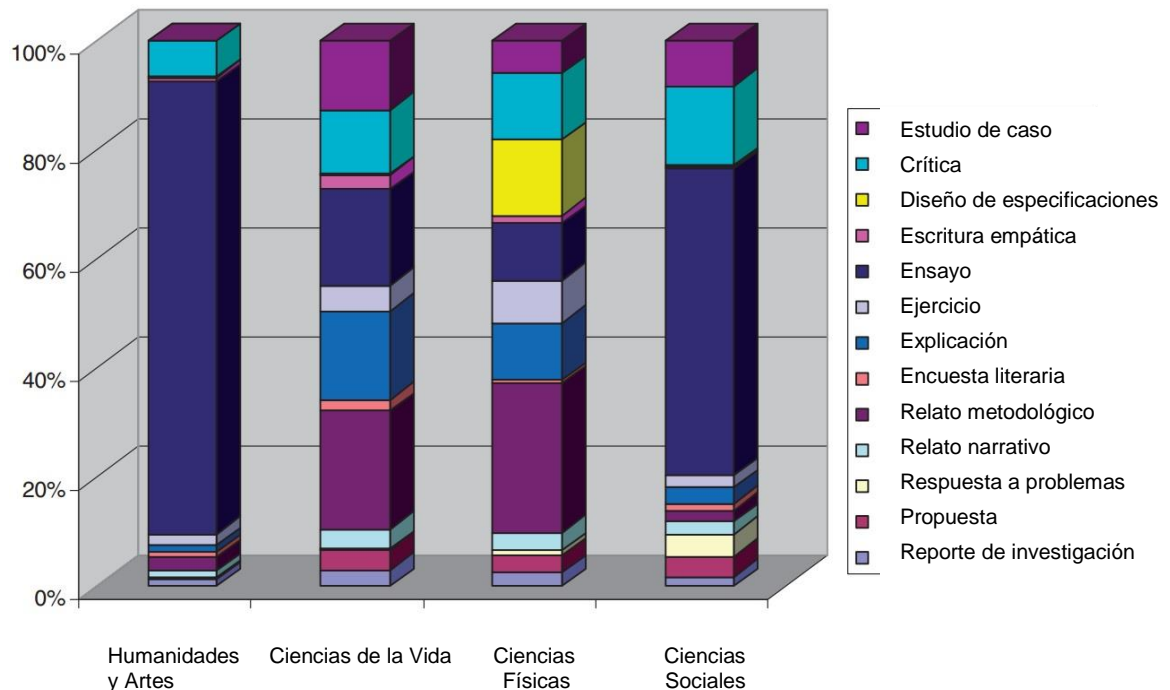
Cuadro 4.2: Tareas de escritura requeridas en el contexto universitario anglófono
(Elaborado por el investigador)

Análisis textual de alumnos	
Autor Método Área Grado	Horowitz (1986) Análisis textual de 54 trabajos escritos 17 programas universitarios Estudiantes de licenciatura
Tipologías textuales identificadas	<ul style="list-style-type: none"> • resumen de lecturas (de artículos o libros) • bibliografía comentada • reporte basado en la experiencia • conexión entre la teoría y los datos • estudio de caso • síntesis de múltiples fuentes • proyecto de investigación
Autor Método Área Grado	Hale et al. (1996) Análisis textual de 188 trabajos escritos 5 disciplinas Estudiantes de licenciatura y posgrado
Tipologías textuales identificadas	<ul style="list-style-type: none"> • ensayos • trabajos monográficos • tareas cortas (menos de media cuartilla) • reportes de experimentos con interpretación • resúmenes • estudios de caso • planes o propuestas • programas documentados de computación • reseña de libros
Autor Método Área Grado	Gardner y Nesi (2012) Análisis textual de 2761 trabajos escritos 30 disciplinas Estudiantes de licenciatura y posgrado
Tipologías textuales identificadas	<ul style="list-style-type: none"> • ensayo • relato metodológico • críticas • estudios de caso • explicaciones • ejercicios

	<ul style="list-style-type: none"> • diseño de especificaciones • relato narrativo • propuestas • reporte de investigación • escritura empática • encuesta literaria • respuesta a problemas
Recolección de datos de cursos académicos	
Autor Método Área Grado	Canseco y Byrd (1989) Sistematización de 55 programas de curso Negocios y administración Estudiantes de posgrado
Tipologías textuales identificadas	<ul style="list-style-type: none"> • exámenes • problemas y tareas • proyectos • artículos • trabajos • estudios de caso • reportes • miscelánea
Autor Método Área Grado	Cooper y Bikowski (2007) Sistematización de 200 programas de curso 10 ciencias sociales y humanidades 10 ciencias exactas Estudiantes de posgrado
Tipologías textuales identificadas	<ul style="list-style-type: none"> • monografías • reportes de experimentos o proyectos • reseñas de artículos o libros • planes o propuestas • resúmenes • estudios de caso • escritura no estructurada • artículos de publicación • ensayos • bibliografía comentada • miscelánea

Entre estos estudios, destaca el de Gardner y Nesi (2012), en el cual las autoras describen trece tipologías textuales diferentes. La Gráfica 4.1 muestra la comparación entre las distintas tipologías textuales pertenecientes a las cuatro áreas del saber. En primer lugar, destacan los ensayos, mismos que corresponden al 83% y al 56% del total de trabajos solicitados a los alumnos de humanidades y artes, y de ciencias sociales, respectivamente. En el caso de las ciencias biológicas, las tipologías textuales: estudios de caso, críticas, ensayos, explicaciones y relatos metodológicos, se encuentran balanceadas de una manera más o menos equitativa. Lo mismo sucede con las ciencias físicas, aunque además de las anteriores, destaca también el diseño de especificaciones.

Gráfica 4.1: Clasificación de familias de géneros textuales
(Tomado de Gardner y Nesi, 2012, p. 46).
(Traducido por el investigador)



Dentro del contexto hispanohablante, Hernández y Castelló (2014) retomaron las tipologías de géneros discursivos descritas por Gardner y Nesi (2012) y condujeron un estudio con alumnos de contaduría de la Universidad Veracruzana en México. Las autoras elaboraron una encuesta en escala tipo Likert para que los estudiantes identificaran la regularidad con la cual escribían estos géneros discursivos. Del estudio, resaltan las autoras que las tipologías textuales recurrentes son: ejercicios, ensayos y relatos metodológicos². Si bien es cierto que una de las limitantes del estudio es que se restringe a una sola disciplina, el estudio representa un primer acercamiento a los géneros discursivos que se emplean en el contexto hispanohablante.

Otra forma de aproximarse a los géneros discursivos de las distintas disciplinas académicas ha sido mediante la Lingüística Sistémico Funcional (LSM). Estos estudios cobraron

² Los relatos metodológicos son propios de las ciencias naturales y experimentales. El propósito de éstos relatos es proveer a los estudiantes de los procedimientos característicos de una disciplina particular. Por ejemplo, en el caso de las ingenierías, se pueden desarrollar reportes de laboratorio, o en química, procedimientos experimentales. En todo caso, estos reportes siguen un patrón específico en cada una de las disciplinas, dependiendo del tipo de estudio del cual se trate.

fuerza en los 80 con las aportaciones de Halliday y Hasan (1980) y han continuado hasta la actualidad con los estudios de Martin (1997) y Martin y Rose (2003, 2008) sobre géneros discursivos. Para estos últimos, el género discursivo tiene una finalidad específica dentro de un contexto situacional, pero también cultural, como sería el caso del texto académico. Uno de los objetivos de la LSM consiste en interpretar los elementos de las estructuras lingüísticas. De ahí que, actualmente se emplee en el ámbito académico para conocer qué aspectos gramaticales, sintácticos y oracionales caracterizan a un género discursivo en particular. En el Estudio 3 de la presente investigación, se describirá cómo se ha empleado la LSM para encontrar características comunes o disímiles entre los textos académicos que se producen en las distintas áreas del saber y aquellos que se generan en los ECLAA.

4.2.1.3. LA IMPORTANCIA DE LA ESCRITURA ACADÉMICA A NIVEL POSGRADO

Un punto crucial dentro de la evaluación de las habilidades de escritura académica radica en la visión de los docentes respecto de su valor dentro del contexto universitario. Sin embargo, aun cuando es cierto que éstos consideran que la escritura es de suma relevancia a nivel posgrado, su posicionamiento ante el fomento y el desarrollo de dicha habilidad es generalmente distante. Delcambre y Donahue (2012) mencionan que es común que los docentes no fomenten dicha práctica entre sus estudiantes o que consideren que no son ellos quienes deberían formar parte del proceso de adquisición de dicha habilidad. De hecho, pese a que es común que los docentes consideren que sus estudiantes no son competentes para emplear de forma eficaz las habilidades de escritura (Chitez y Kruse, 2012), éstos primeros rara vez se reconocen como pieza clave en la adquisición efectiva de dicha habilidad.

Lo que resulta paradójico es que los docentes resalten la importancia de las habilidades de escritura en el ámbito académico y que evidencien carencias en el manejo de éstas por parte de sus estudiantes, y sin embargo, la inclusión de cursos o talleres de escritura en programas curriculares de las universidades hispanoparlantes sea prácticamente nula o de forma aislada (Castelló e Iñesta, 2012). Sin entrar en discusiones sobre si lo mejor sería la inclusión de la escritura académica a través de las disciplinas, de forma curricular o mediante cursos generales de escritura, la realidad es que no se ha atendido la necesidad de su instrucción explícita cualquiera que sea la modalidad. En el caso particular de la UNAM, los cursos de escritura académica rara vez forman parte del currículo y generalmente se encuentran desarticulados de los planes de estudio de las distintas facultades.

La escritura académica representa una herramienta indispensable para poder desarrollar un pensamiento crítico entre los estudiantes. No obstante, dar por sentado que los estudiantes universitarios cuentan con dicha habilidad ha relegado la importancia de incluir cursos de escritura académica a nivel universitario y sobre todo a nivel posgrado. Una de las metas en la educación es justamente la de fomentar la capacidad crítica y argumentativa de los estudiantes; sin embargo, ¿cómo lograr dichas metas si los estudiantes no han explotado al máximo sus competencias comunicativas y particularmente las habilidades de escritura para poder hacerlo? Por consiguiente, si en realidad deseamos que los estudiantes del nivel superior educativo aprendan a construir el conocimiento y a generar propuestas propias a nivel gnoseológico y filosófico, como lo plantea González (2014), paralelamente debemos enseñarles a desarrollar las habilidades de escritura que requerían para transformar y generar conocimiento de forma creativa y auténtica.

A este respecto, Zhizhko (2014) agrega la importancia de contar con una metodología a nivel posgrado que fomente las habilidades de lectoescritura académica en los estudiantes de este ciclo universitario. La autora afirma que para que esto pueda lograrse, deben considerarse, además de la lengua escrita, los distintos géneros discursivos pertenecientes al ámbito científico. No obstante, para que esta propuesta surta el efecto deseado, los cursos de escritura deben encontrarse alineados a los programas curriculares de cada una de las disciplinas; ya que como afirma Flores (2014), los requerimientos de escritura académica no son iguales para un ingeniero o para un filósofo. El dominio efectivo de estos géneros discursivos por parte de los estudiantes de posgrado requiere de la participación interdisciplinaria de especialistas de la lengua y también de los docentes que emplean ciertas tipologías textuales con características propias de la comunidad discursiva a la cual pertenecen.

Finalmente, otra de las problemáticas que se está acrecentando se debe al impulso que se le está otorgando a los estudiantes universitarios, particularmente a los de posgrado, para que escriban en inglés. Entonces, el valor de las habilidades de escritura que debe poseer un estudiante hispanohablante en lengua materna se está dando por *de facto* y el foco está virando hacia la posesión de dichas habilidades en inglés. De hecho, Wächter y Maiworm (citados en Gardner y Nesi, 2012) ya no sólo destacan la importancia de la escritura en inglés en países anglófonos, sino que además apuntan la creciente cifra de universidades europeas pertenecientes a países no angloparlantes que están impartiendo cursos universitarios en inglés. Estos autores, tras un sondeo en 2200 instituciones de educación superior de países europeos

no angloparlantes, identificaron que el 38% de éstas ofrece programas en inglés, de los cuales, la mayoría fueron creados a partir de 2003.

Naturalmente, la influencia de la escritura en inglés ha permeado el contexto hispanohablante, por lo que hoy en día la presión por escribir en inglés con el objeto de poder ser leído por una audiencia internacional ha agravado aún más las dificultades de la escritura académica (Leki, Cumming y Silva, 2008). Razón por la cual, en diversos posgrados de la UNAM, así como de otras universidades, está creciendo vertiginosamente la difusión del conocimiento académico mediante publicaciones en inglés. De hecho, las búsquedas por Internet sobre cursos de escritura académica en la UNAM despliegan en primera posición los cursos de escritura científica en inglés que se están ofreciendo en dicha institución para que los estudiantes y docentes de posgrado publiquen en esta lengua³. De manera opuesta, por ejemplo en el contexto anglófono, instituciones como la Universidad de British Columbia en Canadá ofrece cursos de escritura académica en las siguientes modalidades: escritura académica en inglés y en francés para estudiantes de bachillerato, cursos de gramática en inglés, escritura científica, escritura en la universidad, escritura para estudiantes de posgrado, escritura bajo presión, escritura con estilo, por mencionar los cursos más relevantes⁴.

La situación anterior denota la problemática que están encarando los estudiantes universitarios de la UNAM para desarrollar sus habilidades de escritura en lengua materna al no contar con cursos o talleres especialmente diseñados para el contexto universitario. Paralelamente se vislumbra claramente la dificultad que afrontan al tener que publicar artículos en inglés cuando no cuentan con las herramientas esenciales de escritura académica en español.

En los cursos de escritura académica se requiere de la participación bipartita de los especialistas de la lengua y de los especialistas de los géneros discursivos propios de una comunidad discursiva. No obstante, en los cursos de escritura para publicar en inglés se requiere adicionalmente de especialistas en esta lengua y también de especialistas editores que conocen

³ Consúltese la página del posgrado de la Universidad Nacional Autónoma de México <http://www.posgrado.unam.mx/es/main-menu/escritura-academica-en-ingles>
Última consulta el 26 de diciembre de 2016.

⁴ Consúltese la página de la Universidad de Columbia Británica en la cual el lector podrá encontrar información completa de los cursos que ofrece dicha universidad en escritura académica. <https://cstudies.ubc.ca/study-topic/academic-writing>
Última consulta el 26 de diciembre de 2016.

los requerimientos y procedimientos propios de lo que las editoriales buscan en los artículos que desean publicar. Entonces, ¿si los estudiantes de posgrado no cuentan con las herramientas lingüísticas necesarias para escribir en un contexto académico en lengua materna, en este caso en español, de qué manera podrán responder satisfactoriamente ante la necesidad de hacerlo en inglés? Consecuentemente, los retos que encara la escritura académica en español, no sólo en México, sino en la gran mayoría de los países hispanoparlantes, requiere de estudios que permitan comprender este fenómeno desde diversas artistas. Sin embargo, una de las mayores dificultades se debe a que los pocos estudios que se conducen en materia de escritura académica, por lo menos en el contexto mexicano, tienden a dirigirse hacia la escritura en inglés en programas universitarios bilingües (véase Cuatlapanzi y Perales, 2010; Encinas y Keranen, 2010; Vidal y Perales, 2010).

En este sentido, el análisis del ámbito académico de este estudio pretende dar cuenta de la importancia de la escritura en lengua materna en el contexto hispanohablante a nivel posgrado de la UNAM. En primer lugar, a través del valor que le otorga el docente a la escritura mediante las prácticas que emplea para evaluar el conocimiento de sus estudiantes y en segundo lugar, mediante la jerarquización de las habilidades de escritura que los docentes de dicha institución consideran que son importantes para que los estudiantes sean académicamente competentes.

4.2.2. LAS TAREAS DE ESCRITURA DE LOS ECLAA

Las tareas de escritura que frecuentemente se emplean en los ECLAA han sido duramente criticadas por “carecer de autenticidad respecto del ámbito para el cual pretenden evaluar”. Según los detractores de dicha modalidad (Leki et al., 2008; Moore y Morton, 2005; Uysal, 2010), ningún estudiante universitario escribe un texto en 30 minutos, a mano y sin fuentes previas para construir conocimiento. No obstante, para poder diseñar un examen, se debe circunscribir el universo de generalización a partir del ámbito para el cual deseamos evaluar (Kane, 2006). En el caso de las pruebas escritas de los ECLAA, dicha labor consiste en determinar qué tareas de escritura empleadas en el contexto universitario podrían ser generalizables de este universo. Sin embargo, en el caso del ámbito académico, una de las dificultades reside en lo amplias y diversas que son las prácticas de escritura a nivel universitario. De hecho, Bachman (2002) menciona que aunque las tareas de escritura propuestas para un examen deben estar relacionadas con tareas pertenecientes al mundo real, éstas no necesariamente son reales, aunque sí se busca que sean auténticas.

Lo anterior genera la disyuntiva respecto de establecer un balance entre autenticidad y estandarización (Bachman, 2002; Kane et al., 1999; Weigle, 2002). Estos autores aseveran que en ámbitos que son muy amplios, la incorporación de tareas con alto grado de autenticidad dificulta tanto su corrección como su estandarización, generándose sesgos en los resultados obtenidos y entorpeciendo el proceso de operacionalización de la misma. Los autores comentan que en los exámenes de alto impacto es importante que éstos se encuentren estandarizados, siempre y cuando lo anterior no vaya en detrimento de la autenticidad. Sin embargo, la polémica que se ha desatado en contra de los exámenes estandarizados versa precisamente en la generalización y toma de decisiones que se efectúan a través de un micro universo seleccionado del ámbito en cuestión (Huot, O'Neill y Moore, 2010).

Weigle, por ejemplo, menciona cuatro razones por las cuales simple y llanamente las tareas de escritura de un ECLAA difieren con respecto a un escrito desarrollado en el ámbito universitario:

A typical writing test – in which examinees must write a timed impromptu essay on a topic about which they had no advance knowledge – is a relatively inauthentic in at least four ways. First, academic writing outside of a testing situation typically involves the use of source materials as input... Second, except in the case of timed examinations, most academic writing is not timed or speeded, and writers can take as much time... and revise their writing before turning it in for a grade. Third, while writing tests are frequently scored by raters who are unknown to the examinees (and vice versa), the audience for most academic writing is the students' instructor... Finally, most academic writing... is judged primarily on accuracy of content, rather than on the appropriateness of the organization or the use of language (Weigle, 2002, p. 52).

No obstante, pese a que las tareas propuestas para un examen no sean tareas reales que un estudiante enfrentará al momento de cursar estudios universitarios, eso no significa que carezcan de autenticidad. Bachman (2002) asevera que uno de los objetivos de la selección de tareas consiste en la generalizabilidad; por lo tanto, entre más “real” sea una tarea, menos generalizable será y mayores riesgos correremos al establecer inferencias de lo que el sustentante es capaz de hacer. Sin embargo, también es importante delimitar el alcance de una tarea en cuanto a las habilidades de escritura que un examinado demuestra para resolver la tarea, puesto que como menciona Bachman (2002), predecir futuros desempeños resulta riesgoso.

Entonces, aunque Weigle (2002) menciona que las tareas propuestas para los exámenes de lengua no son reales, esto no significa que sean inauténticas. La diferencia fundamental

estriba en que se pretende evaluar los conocimientos lingüísticos de un sustentante y la manera mediante la cual hace uso de ciertas habilidades y estrategias de escritura que son frecuentemente empleadas en el ámbito académico; por ejemplo, como la capacidad de argumentar. Consecuentemente, en contraposición a lo que afirma Weigle (2002), una tarea de examen puede ser auténtica, pero no significa que sea real dentro del contexto para el cual se evalúa.

Más aún, Bachman y Palmer (1996, citados en Bachman, 2002, p. 460) sugieren tres razones por las cuales las tareas del “mundo real” no son apropiadas en ámbitos que son muy amplios: En primer lugar, no todas las tareas relacionarán las habilidades que deseamos evaluar. En segundo lugar, algunas tareas son imprácticas para administrarse en un examen. Y en tercer lugar, no todas las tareas reales serán apropiadas para todos los sustentantes, sobre todo si éstas presuponen conocimiento previo del tema o del desarrollo específico de la tarea. Mendoza (2014) cuestiona cómo poder elegir entre una de las diferentes tipologías textuales comúnmente empleadas a nivel universitario cuando resulta difícil sistematizar los géneros discursivos y las tareas de escritura que les son solicitadas a los estudiantes universitarios.

En aras de lograr mayor autenticidad en el diseño de los ECLAA, la inclusión de tareas de escritura (conocidas también como pruebas directas) ha aumentado considerablemente y además ha desplazado en gran medida las prácticas de evaluación de las habilidades de escritura mediante modelos psicométricos (conocidos como pruebas indirectas). Ésta última, generalmente bajo la modalidad de opción múltiple, fue ampliamente utilizada durante la segunda mitad del siglo XX debido a su practicidad y facilidad de corrección. Sin embargo, el advenimiento del enfoque comunicativo de los 80 (Canale y Swain, 1980 y Canale, 1983) favoreció cambios importantes en el diseño de exámenes de lengua. Más adelante, en los 90, la necesidad de diseñar exámenes que reflejaran la realidad y las condiciones de uso en contextos específicos (Bachman y Palmer, 1996) aceleró el florecimiento de dichas tareas de escritura.

En el caso del IELTS (conocido hasta 1989 como ELTS), por ejemplo, se incluyó en 1980 una tarea de escritura integrada en la cual los candidatos debían remitirse a uno de los textos, previamente leídos en la sección de comprensión de lectura, para componer su redacción. Tras diversas investigaciones conducidas con el fin de validar la prueba, entre las que destacan la de Alderson y Clapham (1993), en 1995 se decidió eliminar la tarea integrada debido al sesgo en la evaluación que se presentaba al mezclar la habilidad de lectura con la producción escrita, y se

optó por una prueba escrita compuesta por dos tareas discretas (una con información visual –generalmente a través de gráficas– y otra con una consigna que requiere de la argumentación u opinión para su elaboración).

La contraparte norteamericana del IELTS, el TOEFL PBT (*Paper-Based Test*), adoptó por vez primera en 1986 una tarea de escritura a modo de ensayo. Sin embargo, dicha prueba era independiente del examen y solamente se administraba en casos en los cuales era solicitada por las instituciones. La prueba escrita del TOEFL se convirtió en obligatoria hasta 1998 con el nacimiento del TOEFL CBT (*Computer-Based Test*). No obstante, este examen no fructiferó y rápidamente fue remplazado por la nueva versión denominada TOEFL iBT en 2005. Esta nueva versión incluye una tarea de escritura integrada con comprensión de lectura y comprensión auditiva, y una tarea discreta a manera de ensayo. Sin embargo, la prueba integrada ha sido duramente criticada debido a su extensión, puesto que el supuesto trasfondo de que un estudiante es capaz de integrar el material proveniente de dos fuentes (una oral y otra escrita) y condensarlo en un texto de 150 palabras es poco plausible de su capacidad para construir conocimiento a nivel universitario. En este sentido, como bien mencionan Kane et al. (1999), establecer el balance entre estandarización y autenticidad no es fácil. En el caso del contexto hispanohablante, no existe aún ningún ECLAA que sea administrado y reconocido a nivel nacional o internacional.

4.2.2.1 LOS TIPOS DE TAREAS DE ESCRITURA PARA LOS ECLAA

En tan sólo un par de décadas, el salto de las pruebas de escritura indirectas a directas ha conllevado un reto para quienes se dedican a diseñar y validar exámenes. La selección de las tareas y las variables intrínsecas a que éstas están sujetas han suscitado una serie de controversias respecto de su autenticidad y pertinencia, pero sobre todo, ha generado cuestionamientos sobre su confiabilidad, validez y, por consiguiente, sobre la equidad y las consecuencias benéficas en la evaluación. Por ende, tanto en el diseño como en la evaluación de las pruebas de escritura de respuesta construida, resulta indispensable sopesar las diversas dificultades que giran en torno a las variables que inciden favorable o desfavorablemente en éstas con el fin de dilucidar qué tareas resultan óptimas para evaluar la habilidad escrita. Weigle (2002), en su libro sobre evaluación de escritura: *Assessing Writing*, hace una extensa descripción y evaluación de las dificultades que encara la evaluación de la escritura en exámenes de lengua, particularmente aquella destinada a los ECLAA, pues es la que mayormente se ha sistematizado y analizado. Estas problemáticas serán descritas con detalle más adelante.

Debido a la problemática que representa la concepción, diseño y selección de tareas de escritura para evaluar las habilidades de escritura en un contexto académico, es necesario diseñarlas y pilotearlas adecuadamente con el fin de determinar su pertinencia antes de emplearse como forma de evaluación. De hecho, la selección de la tarea y la variedad de muestras de un mismo tipo de tareas representa una de las fuentes de variabilidad importantes a considerar dentro de la validez de un examen (Barkaoui, 2007; Jonsson y Svingby, 2007).

Dado que la PE cuenta con sólo una o dos tareas, se espera que los candidatos exploten al máximo sus recursos y habilidades de escritura. Weigle (2002) sugiere que en la etapa del diseño se obtenga la mayor información referente al propósito del examen, las características de la población meta, las necesidades del “mundo real” y los recursos con los que se cuentan para administrarla y para evaluarla. En cuanto al diseño del examen, resulta indispensable hacer un análisis de necesidades del ámbito para el cual deseamos evaluar, así como de las habilidades que deseamos evidenciar mediante la prueba, pues como menciona la autora, las demandas de escritura académica conllevan mayor destreza en el uso de la lengua para alcanzar determinadas funciones comunicativas que requieren de conocimiento funcional y sociolingüístico especializado con miras a persuadir a una audiencia académica. De hecho, Bachman (2002) agrega que en la etapa del diseño se debe describir claramente las especificaciones de lo que se pretende evaluar mediante la expresión escrita, por ejemplo, en aras de posteriormente poder seleccionar tareas que sean auténticas.

Weigle (2002) añade que las consideraciones más importantes que deben tomarse en cuenta para el diseño de tareas escritas son: a) la descripción del ámbito de escritura, b) diseñar consignas claras y precisas que sean interpretadas por los sustentantes de la misma forma; es decir, sin ambigüedades ni dificultades léxicas o sintácticas que entorpezcan la comprensión de la tarea y c) uniformidad en la forma de evaluación, con lo cual se puedan obtener resultados consistentes en evaluaciones posteriores o al ser revisados por diferentes correctores. Debido a que la prueba escrita de un ECLAA cuenta con tan sólo una o dos tareas a desarrollar en aproximadamente una hora, la cuidadosa selección de la tarea, su correcta construcción y la enunciación de una consigna clara, resultan pasos fundamentales para que éstas cuenten con el nivel de dificultad deseado y así, evitar sesgos en los resultados obtenidos por los examinandos. Por esto, con el fin de esclarecer algunos de los cuestionamientos más controversiales en relación con las tareas de escritura para los ECLAA, se analizarán las

variables más representativas que juegan un papel importante en el desarrollo de las tareas de escritura.

4.2.2.2. LAS VARIABLES INTRÍNSECAS ASOCIADAS AL TIPO DE TAREA ESCRITA

Las variables que influyen en el desempeño de la tarea son múltiples. Bachman (2002) y Bachman y Palmer (1996, p. 65) mencionan que al momento de resolver una tarea entran en juego varios factores como los personales (edad, lengua materna, nivel de educación, familiarización con el examen, conocimiento del mundo, conocimiento de la lengua) y los contextuales, generalmente determinados por las características de la tarea. Asimismo, afirman que hay un proceso de interacción en el cual se ven involucrados esquemas socioafectivos y también el uso de estrategias metacognitivas para su resolución. Dado que el objetivo de esta investigación consiste simplemente en ver qué tan pertinentes son las tareas de escritura seleccionadas respecto del ámbito para el cual se desea evaluar, solamente se abordarán las variables relacionadas con la tarea. A este respecto, Weigle (2002) menciona que las variables más representativas que inciden en el diseño de las tareas de escritura son: la posibilidad de elegir entre varias tareas, el contenido, el conocimiento previo del tema, el uso de material de estímulo, el modo discursivo –mismo que incluye el género textual, la retórica, el patrón de exposición y las demandas cognitivas–, la longitud del texto, el tiempo de respuesta, la consigna, el modo de escritura y el uso de diccionarios.

4.2.2.3. LA SELECCIÓN DE LA TAREA

La decisión más importante consiste en seleccionar la tarea o tareas a través de las cuales se evaluará la habilidad escrita. La razón fundamental reside en que el tipo de tarea, dependiendo del fin para el cual se desee evaluar, despliega una serie de variables que en mayor o menor medida facilitan o dificultan la resolución de la misma, poniendo así en cuestionamiento la equidad y la evaluación justa de los candidatos que son examinados. En cuanto a la dificultad de la tarea, Bachman (2002) afirma que ésta no reside en la tarea per se, sino que más bien se encuentra determinada por el proceso de interacción que se establece entre el sustentante y la tarea. Entonces, quizás lo que habría que considerar al momento de diseñar tareas de escritura es que éstas fomenten la interacción y puedan establecer eficazmente un vínculo socioafectivo entre el sustentante y la tarea.

En general, podríamos decir que las tareas de escritura de los ECLAA se dividen en dos: discretas (miden una sola habilidad) e integradas (relacionadas alguna otra habilidad). La

diferencia fundamental estriba en que en las primeras solamente se evalúa la expresión escrita a partir de un estímulo dado por la consigna, una frase o una imagen, mientras que en la segunda el examinado debe procesar primero input oral o escrito para después escribir. Ejemplos de tareas discretas serían: escribir a partir de un estímulo visual (gráficas o esquemas de procedimiento) o escrito (consignas para redactar ensayos o tareas generativas y refranes o frases para buscar su interpretación y postura). En el caso de las tareas integradas, podemos encontrar aquellas que incluyen información escrita y/u oral. Aunque la selección de una tarea de escritura pareciera una decisión sencilla de tomar, en realidad no lo es, ya que la crítica que ha recibido toda esta amplia gama de posibilidades reside en que cada una de ellas involucra variables adicionales que interfieren directamente con el nivel de dificultad de la tarea; generándose así posibles sesgos en los resultados obtenidos.

4.2.2.3.1. LAS TAREAS DISCRETAS

Debido a que las tareas discretas involucran la redacción de un tema a partir de una consigna dada, la mayor controversia emerge al seleccionar tareas que involucran o no el conocimiento previo del sustentante. A este respecto, hay quienes abogan por tareas derivadas de temas académicos relacionados con el área de especialización o conocimiento del candidato, mientras que por otro lado, hay quienes arguyen que lo anterior pone en ventaja a quienes conocen más sobre un tema en particular, lo cual generará sesgos en los resultados obtenidos y, consecuentemente, menor confiabilidad en la prueba (He y Shi, 2012). Estos últimos autores retoman a Carlson y Bridgeman (1986, p. 139), quienes apuntan que “el contenido dado por el tema debe ser tan justo como sea posible, sin que éste favorezca a un grupo específico ni a experiencias culturales”. Esta desventaja se incrementa aún más cuando se trata de estudiantes de posgrado (Hamp-Lyons, 1991b; Lee y Anderson, 2007; Weigle, 2002). Lee y Anderson (2007) añaden que el riesgo que subyace al diseñar tareas de escritura de acuerdo al área de estudio de los candidatos se debe a que hay áreas que son interdisciplinarias por naturaleza, lo cual dificultaría aún más la ardua tarea de sintonizarlas de acuerdo con campos del conocimiento que no sólo son numerosos, sino además heterogéneos y multidisciplinarios. Aunado a lo anterior, con base en el estudio que los autores llevaron a cabo –en el cual son asignados tres temas aleatoriamente a los estudiantes– debido a que la dificultad variaba de tema a tema, lo más recomendable sería asignar un tema a toda la misma población.

Sin embargo, en otro estudio que Lee (2008) condujo con dos tareas, una con un tema general y otro específico con cuatro opciones a elegir pertenecientes al área de negocios,

humanidades, tecnología y ciencias de la vida, no encontró diferencias significativas en los resultados de ambas pruebas. La diferencia fundamental entre ambos estudios estriba en que en el primero se le da al sustentante un tema específico de forma aleatoria, mientras que en el segundo, el tema pertenece al área en el cual se encuentra el sustentante. No obstante, si bien es cierto que Lee (2008) identificó que no había diferencias sustanciales en los resultados de los estudiantes que escribieron sobre un tema general y otro relacionado con su área general de estudio, el autor sugiere que por practicidad sería mejor elegir un tema general, pues la otra opción podría afectar la manera en que los evaluadores califican.

Naturalmente, el tema seleccionado y la consigna determinan el uso o no de conocimiento previo y éste, a su vez, impacta en el contenido. Weigle (2002) comenta que debido a las múltiples variables relacionadas con el candidato (antecedente académico, estilo de aprendizaje, edad, sexo, lengua materna, conocimiento sobre el tema, grado académico y tiempo de residencia en el país) lo mejor será excluir el conocimiento de un tema en particular, razón por la cual las pruebas que no evalúan el conocimiento previo encajan bien en los exámenes de alto impacto. Sin embargo, lo anterior ha suscitado controversias en cuanto a la selección de la tarea; puesto que dentro de la escritura académica universitaria se busca que éstos no sólo lo hagan de manera coherente y siguiendo las normas convencionales de escritura, sino que además sea de manera responsable: citando a las fuentes y evitando el plagio. Por ello, Leki et al. (2008) agregan que una de las críticas más fuertes contra los ECLAA es que las tareas que generalmente se plantean eliden completamente el uso de fuentes externas.

Otro tipo de tarea discreta (aunque también se podría trabajar con la modalidad integrada) recientemente implementada es la denominada generativa. Su objetivo primordial es incitar la metarreflexión de los estudiantes respecto de su aprendizaje o propósitos a futuro (Condon, 2009; Petersen, 2009). Petersen condujo un estudio en el cual identificó que de un total de 893 estudiantes que respondieron un ensayo de tipo argumentativo, 36 de ellos (4%) mostraron oposición y resistencia abiertamente con frases y títulos que explícitamente retaban el tema para redactar. Condon (2009), por su parte, menciona que debido a la dificultad que subyace al elegir entre tareas de tipo general y aquellas destinadas a áreas específicas del saber, se debería optar por aquellas consideradas de tipo generativo; es decir, aquellas que le permitan al estudiante analizar sus propias experiencias como escritores dentro de un contexto académico o como estudiantes. El autor comenta que este tipo de tareas presenta varias ventajas; por ejemplo, el interés del sustentante en escribir sobre un tema que lo relaciona directamente con su vida

académica estimula el proceso metarreflexivo en el estudiante y, agrega, más allá de la evaluación como un proceso mecanizado que solamente busca ubicar a los estudiantes en un punto específico dentro de una escala, se genera un corpus que permite ahondar en investigaciones respecto de los diferentes procesos de aprendizaje.

Generative prompts – whether about learning goals or other topics that matter to test-takers as well as their institutions – reverse that emphasis, producing writing that matters, first to the test-takers, then to their institutions, and finally to researchers from a broad variety of fields... Once a prompt elicits writing that is interesting and useful, then the timed writing itself becomes a student learning outcome (Condon, 2009, p. 152).

Contrariamente a lo que menciona Condon (2009), valdría la pena matizar que quizás la oposición abierta de escribir sobre un tema no reside en la dificultad de la tarea, sino más bien en la falta de elementos en la consigna que estimulen el esquema socioafectivo del estudiante para interactuar apropiadamente con la tarea. En este sentido, se podría argumentar que no es la tarea generativa por sí misma la que estimula al estudiante a escribir, sino la propuesta de incluir temas que generen un vínculo con el estudiante y, por lo tanto, favorezcan el proceso de interacción. Por otro lado, habría que considerar el uso de dichas tareas en poblaciones extranjeras provenientes de sistemas escolares diversos y que no cuentan con los mismos antecedentes académicos similares, dado que fomentar el proceso reflexivo como lo plantea Condon (2009) generaría sesgos importantes. Sin embargo, lo que sí vale la pena rescatar de las tareas generativas es que éstas buscan una estrecha relación con el sustentante y con su experiencia. Kormos (2012, p. 394) agrega que “los estudiantes que se encuentran intrínsecamente motivados se involucran en el proceso de aprendizaje porque encuentran el proceso de escritura interesante y placentero, mientras que los estudiantes extrínsecamente motivados realizan una actividad de aprendizaje con el fin de obtener una recompensa o evitar un castigo”. Por lo tanto, las tareas deberían estimular el factor motivacional del examinando tanto en contextos académicos como en exámenes de dominio para el ámbito académico. A este respecto, Lee (2008) menciona que los estudiantes prefieren elegir entre varias temáticas, pues además de incrementar su motivación, los estudiantes buscan temas con los cuales tengan mayor familiaridad y destreza para escribir.

La última variante de tareas discretas es aquella en la cual el estímulo que se proporciona para escribir es visual. Este tipo de información generalmente se da a través de gráficas para describir y contrastar información, o de imágenes que denotan cómo se realiza un procedimiento. En estos casos, el estímulo visual presenta el contenido per se, lo cual permite que el candidato recurra a su conocimiento lingüístico y léxico para describirlo. Weigle (2002) arguye que la

carencia de estímulo escrito u oral permite que el candidato explote al máximo su conocimiento del mundo y sus recursos lingüísticos. Sin embargo, al seleccionar este tipo de tareas se debe tener la cautela de escoger elementos visuales que sean lo suficientemente comprensibles y claros, así como con contenido que no sea especializado, ya que esto pondría en desventaja a quienes carecieran del conocimiento léxico sobre el tema en cuestión.

Entonces, en el caso de las tareas discretas que incluyen gráficas se evita, por una parte, que el estudiante copie literalmente paquetes léxicos, mientras que por otra parte, el material visual delimita la tarea y contiene toda la información que el estudiante necesita para transformarla en discurso escrito. Sin embargo, Yang (2012) apunta que las tareas integradas con material visual estimulan el uso de determinado léxico que representa un reto mayor que la dificultad sintáctica. Uysal (2010), en la revisión crítica que hace del IELTS académico, el cual incorpora una gráfica para describir, comenta que las tareas más simples con menos información visual permiten un mejor desempeño de los candidatos; por lo cual, la selección de las tareas que incluyen gráficas deberá ser cuidadosamente seleccionada para evitar sesgos en los resultados. Aunado a lo anterior, la familiaridad con la descripción de gráficas podría constituir un factor de irrelevancia del constructo (Messick, 1989).

Shah y Freedman (2009) comentan que las gráficas presentan diversas variables que inciden dentro de la tarea; a saber: el formato de la gráfica (barras, pastel, lineal), la familiaridad que puedan tener los sustentantes con las gráficas y/o con el tema a describir. El estudio de estos autores resulta de gran utilidad para identificar aquellas dificultades que encaran los estudiantes al momento de describir gráficas. Más allá del aspecto lingüístico, los autores buscan identificar de qué manera el formato de la gráfica facilita o dificulta la tarea, además del conocimiento previo, tanto del tema, como de la elaboración de gráficas y la habilidad para describirlas. Sin entrar en detalle respecto de las estrategias que se requieren para describir gráficas, los autores mencionan que el estar familiarizado con los datos permite establecer mejores inferencias, además de que quienes poseen habilidades desarrolladas con la descripción de gráficas pueden interpretarlas mejor. Aunado a lo anterior, el formato de las gráficas lineales dificulta la lectura de los estudiantes menos familiarizados con éstas, mientras que las de barras resultan ser más claras y fáciles de manejar. Todos los factores antes mencionados van más allá de las habilidades de escritura que posee un sustentante, razón por la cual deben ser cuidadosamente controladas por quienes diseñan estos exámenes.

Según Yang (2012, p. 176), la incorporación de tareas integradas con material visual despliega ciertos procesos que deben ser atendidos:

1. Etapa de comprensión, donde la información global y local debe ser identificada y organizada para ser procesada posteriormente.
2. Etapa de interpretación, donde la información clave es seleccionada y las ideas relevantes son conectadas para elaborar el escrito.
3. Etapa de traducción o de composición, donde la información es decodificada o representada en discurso escrito a través de una cuidadosa planeación, evaluación, monitoreo y revisión.

Por último, Yang (2012) concluye que el uso de estrategias es primordial para la consecución de la tarea, por lo cual debería hacerse énfasis en los cursos de preparación la inclusión de este tipo de tareas discretas, así como de las estrategias requeridas para abordarlas. Finalmente, Shah y Freedman (2009) comentan que la importancia de trabajar con gráficas en el ámbito académico es esencial. Los autores reportan que las gráficas empleadas en libros de texto y en revistas científicas se duplicaron de 1984 a 1994 (2009, p. 2). Asimismo, Hyland (2009) señala que pese a las dificultades que puedan enfrentar los estudiantes con respecto a la interpretación de gráficas, la habilidad para analizarlas e interpretarlas es importante para el éxito académico de los estudiantes universitarios.

4.2.2.3.2. LAS TAREAS INTEGRADAS

Aun cuando en la PE del EXELEAA se seleccionaron exclusivamente dos tareas discretas, en este apartado se describirán las tareas integradas, debido a la importancia que juega dicha forma de escritura en el ámbito académico. Las tareas integradas consisten en escribir un texto a partir de información escrita y/u oral que se da como input. Este tipo de tareas han sido incluidas recientemente en los ECLAA con miras a dar cuenta sobre la manera mediante la cual los candidatos construyen el conocimiento a partir de la inclusión de información proveniente de diversas fuentes; por lo tanto, resultan “más auténticas” para el contexto para el cual se desea evaluar: el académico. El supuesto trasfondo reside en que las tareas que los estudiantes realizan en la universidad incorporan la comprensión de lectura y oral; mientras que las tareas independientes se encuentran aisladas de estas habilidades. Plakans (2010) menciona que una cuestión fundamental para la resolución de una tarea escrita reside en la representación que el alumno se hace dependiendo del tipo de tarea que requiere resolver. Entonces, cuando un estudiante identifica la clase de producto que se espera de la escritura, éste atraviesa por un

proceso de entendimiento en el cual puede identificar el tipo de habilidades y procesos que requerirá para llegar a este producto determinado: el tipo de tarea escrita solicitada.

Tras casi una década de la inclusión de una tarea integrada en el examen del TOEFL iBT y tras numerosos estudios que se han efectuado con diversas variables que influyen en la resolución de la tarea, Cumming (2013) establece un balance entre los pros y los contras de la adopción de dicha modalidad, así como de las investigaciones que se han conducido en este sector. El autor apunta que las bondades que entraña dicha tarea son: tareas más auténticas, escritura con contenido “responsable” y evaluación de las habilidades lingüísticas en la construcción-integración del conocimiento. El autor menciona que uno de los objetivos de las tareas integradas consiste en evidenciar la capacidad de los estudiantes de incorporar el contenido proveniente de diversas fuentes. Las desventajas, sin embargo, son: mezcla de otras habilidades en la medición de la escritura, involucramiento de géneros que no están bien definidos y que son difíciles de evaluar, requerimiento de ciertas habilidades mínimas para llevarla a cabo, y la inclusión de información que es difícil determinar si su procedencia subyace en la capacidad de producción del sustentante (p. 2).

A continuación, se describen algunos estudios que han tratado de aportar evidencia sobre el funcionamiento de las tareas integradas, con lo cual se les pueda dar cabida en lugar de las discretas. En la última década se han desarrollado tareas integradas que incorporan material escrito y/u oral de entrada para desarrollar la tarea. Sin embargo, dentro de las dificultades que encarna la implementación de tareas integradas se halla precisamente la incorporación de habilidades orales y de lectura que, en caso de no estar bien consolidadas, dificultarán la producción escrita de los estudiantes (Leki et al., 2008). De hecho, en el estudio que Plakans (2010) condujo para ver de qué forma la representación de la tarea (discreta o integrada) impactaba en el proceso de escritura de los examinados, la autora comenta que la incorporación de textos en las tareas integradas puede servir de apoyo para algunos estudiantes, aunque otros pueden verse desfavorecidos por la complejidad del tópico o por la dificultad que entraña la construcción del conocimiento al incorporar el material de lectura en el escrito solicitado. Aunado a lo anterior, una dificultad más se encuentra en determinar qué fue copiado literalmente del texto, tarea que les resulta sumamente compleja a los evaluadores humanos (Cumming, 2013; Cumming, Kantor, Baba, Erdosy, Eouanzoui y James, 2005).

A este respecto, Weigle y Parker (2012, p. 128) identificaron que el porcentaje tomado literalmente del texto era en la mayoría de los casos menor al 8%, además de que éste se refería normalmente a colocaciones. Sin embargo, se evidenció que los estudiantes con puntajes bajos tienden a citar porciones más largas provenientes del texto de entrada, a diferencia de los estudiantes con mayor dominio de la lengua, quienes parafrasean más. Haciendo eco al estudio de Weigle y Parker (2012), el de Plakans y Gebril (2013) confirmó que el uso apropiado de las fuentes incide positivamente en la escritura. Los investigadores compararon el resultado del TOEFL iBT con tres aspectos importantes del contenido de la tarea integrada; a saber: la importancia de las ideas incluidas en el resumen, el uso de ideas provenientes de la lectura y del audio, y la copia textual de fragmentos del texto. Mediante un análisis de regresión lineal, los autores encontraron que el contenido del audio y la inclusión de ideas importantes provenientes de ambos textos explican en mejor grado una correlación positiva con el puntaje obtenido en la prueba. En el caso del calco literal del texto escrito, los autores encontraron una correlación negativa, lo cual comprueba que los sustentantes con mayor dominio tienden a parafrasear más y a copiar menos directamente del texto. Plakans y Gebril (2013) identificaron que la tarea integrada favorece la inclusión del contenido del audio, mientras que el del texto se emplea solamente para contrastar la información escuchada.

No obstante, las tareas integradas del TOEFL iBT que incluyen un texto de entrada escrito y otro auditivo nos llevan al cuestionamiento de si en realidad la comprensión auditiva, por lo menos dentro del contexto universitario, conlleva tanto peso al momento de escribir trabajos académicos. Por lo tanto, sería importante indagar si las prácticas de escritura universitaria buscan la integración de estas dos habilidades; de lo contrario, valdría la pena considerar si una tarea integrada con dos o múltiples fuentes escritas sería más apropiada que una que incluye comprensión de lectura y auditiva. Siguiendo la línea de Weigle y Parker (2012), ¿cómo distinguir del material tomado del texto cuando éste no ha sido citado? Aunado a lo anterior, ¿se puede evidenciar con un par de textos y un corpus de 150 palabras la habilidad de integrar el material proveniente de diversas fuentes? Como asevera Hyland (2009), el correcto uso del material escrito con su correcta acreditación al autor es un componente importante de la escritura académica –misma que debería figurar dentro del constructo de los exámenes–; sin embargo, la gran interrogante que subyace es si en verdad esta habilidad puede evaluarse en un texto de 150 palabras con una o dos fuentes de entrada.

Finalmente, dentro del contexto hispanohablante, parte del material que se emplea para construir el conocimiento escrito procede de fuentes como el inglés. Por lo tanto, valdría la pena indagar tres cosas: en primer lugar, ¿qué porcentaje de las lecturas solicitadas por los docentes universitarios provienen de lenguas extranjeras o indígenas y cuáles? En segundo lugar, ¿qué porcentaje de las lecturas que hacen los alumnos es en otra lengua diferente del español? Y por último, ¿hasta qué punto sería óptimo emplear tareas integradas con textos escritos en español en disciplinas en donde casi el 100% del material de lectura está en inglés? No obstante, pese a las dificultades y retos que representa el uso de tareas integradas, la incorporación de éstas en exámenes de lengua con fines académicos merece atención especial.

4.2.2.4. OTRAS VARIABLES IMPORTANTES

Existen otras variables que inciden indirectamente en la prueba como: la claridad en la consigna, la planeación de la tarea, el tiempo de respuesta, el modo de escritura y el uso del diccionario. No obstante, pese a que los hallazgos respecto de estas variables no han logrado dilucidar cuál sería la mejor manera abordar dichas variables, resulta importante considerar su relevancia.

4.2.2.4.1. EL MODO DISCURSIVO

Tras haber revisado la variable fundamental que subyace a las tareas de escritura (discretas versus integradas), cabe mencionar que la selección de cualquier modalidad contenida dentro de éstas dos desplegará otra variable importante: el modo discursivo relacionado con la tarea que ha sido elegida. Weigle (2002) comenta que el modo discursivo es particularmente relevante en la evaluación de la escritura de segundas lenguas, pues la manera de desarrollar incluso un mismo género textual varía de cultura en cultura. Por ejemplo, Weigle (2002) comenta que los evaluadores tienden a premiar inconscientemente aquellos modos discursivos que resultan más complejos de elaborar, como el argumentativo, a diferencia de otros que son catalogados como más sencillos; por ejemplo, el narrativo o descriptivo.

El objetivo de la presente investigación dista de un análisis a detalle de todas las posibles variables que inciden en cuanto a la tipología textual, por lo cual se presenta un esbozo de los hallazgos más representativos en cuanto al modo discursivo. Leki et al. (2008) hacen un minucioso recuento de los rasgos más significativos de los patrones de organización retórica de diversos grupos lingüísticos, así como de los modos discursivos que emplean éstos en sus lenguas maternas. Con respecto a la organización retórica, los autores mencionan que ésta

tiende a ser normalmente lineal en los anglosajones y con un esquema que parte de lo general a lo particular; sin embargo, los japoneses, por ejemplo, emplean un acercamiento lineal y circular en sus composiciones tanto en inglés como en japonés, pese a que los patrones retóricos del inglés son más bien de tipo lineal, pero la influencia de la lengua materna se denota en la estructuración retórica. Asimismo, debido a patrones culturales, los japoneses (en comparación con los estadounidenses) tienden a reservarse más sus opiniones o críticas dentro del texto. La retórica de los chinos, agregan los autores, es de tipo “centrifigural interoracional”, puesto que el patrón de escritura de quienes comparten esta lengua consiste en añadir información a una idea inicial y después una serie de ejemplificaciones. En el caso de los hispanoparlantes, el factor emotivo afecta la calidad de los textos en inglés, debido a que la inserción de temas que conllevan una carga emotiva evidencia la irrelevancia en las introducciones y la falta de transiciones con párrafos fuera de foco y una progresión de ideas poco clara (p. 147).

Algunos detractores de la autenticidad de las pruebas escritas de los exámenes con fines académicos (Moore y Morton, 2005, Uysal, 2010) arguyen que éstas son diferentes a las tareas que se realizan en un contexto real. Sin embargo, Bachman (2002) menciona que la autenticidad no significa que una tarea sea real, sino que ésta se encuentre relacionada con el uso de la lengua requerido en el contexto para el cual se desea evaluar; en este caso, el académico. Por esta razón, resulta natural que los ensayos que se redactan en un ECLAA sean diferentes a los ensayos que los estudiantes redactan en su vida universitaria.

Efectivamente, no se puede comprar una tarea de un examen redactada en 30 minutos con un tema predominantemente del ámbito público, sin uso de correctores de texto ni diccionario y sin fuentes ni referencias externas con una tarea que se desarrolla en un contexto académico con numerosas fuentes escritas, recursos electrónicos, correctores de texto, diccionarios, tesauros, conocimiento del tema y tiempo relativamente flexible para su elaboración y edición. Lo que sí se puede evidenciar en una PE de un ECLAA son las habilidades de escritura en lengua meta que el estudiante pone en marcha para construir su texto. De igual manera, estas mismas habilidades de escritura son empleadas en un contexto académico, aunque con recursos y estrategias adicionales que valdría la pena indagar.

En suma, independientemente de la tarea que se elija: discreta o integrada, cada una de ella presenta pros y contras que deben ser tomados en cuenta al momento de diseñar la prueba. Pero más a fondo, valdría la pena considerar tareas que promuevan el proceso de interacción

entre el sustentante y éstas, puesto que sólo de esta manera el candidato podrá establecer un vínculo socioafectivo que le confiera significancia al proceso de escritura y así, podrá desplegar eficazmente estrategias cognitivas y metacognitivas que le permitan explotar al máximo sus habilidades lingüísticas.

4.2.2.4.2. LA CLARIDAD EN LA CONSIGNA

Por ejemplo, en el caso de la consigna, como bien sugiere Weigle (2002) ésta debe ser clara y concisa, puesto que las consignas muy elaboradas que en ocasiones incluyen ejemplos son sujetas a ser copiadas por los estudiantes. Por el contrario, cuando la consigna no es muy clara o demasiado abierta a interpretación, no permite delimitar la tarea y los estudiantes tienden a escribir sobre una vasta variedad de tópicos. En estudios que se han conducido para identificar las secuencias formulaicas o paquetes léxicos se ha evidenciado que los estudiantes menos cualificados usan una mayor cantidad de estas secuencias (Staples, Egbert, Biber y McClair, 2013) y también copian más paquetes léxicos provenientes de las consignas o, en el caso de las tareas integradas, de los textos.

Kobrin, Deng y Shaw (2011) mencionan que la redacción de la consigna determina la dificultad de la tarea. Dicha variable se percibe, por ejemplo, cuando hay palabras o calificadores de una afirmación, cuando se expresa un punto de vista neutral o con una opinión, así como con argumentos con una o dos posturas. Incluso, una consigna demasiado larga podría afectar el entendimiento de aquellos estudiantes con menos dominio lingüístico. No obstante, retomando el concepto de Bachman (2002) sobre la dificultad de la tarea, aquella no es intrínseca a ésta sino más bien a la manera mediante la cual el individuo interacciona con ésta. De tal suerte que más allá de la manera mediante la cual se encuentra formulada la consigna y el tema, la relación de éste con el sustentante determinará la manera mediante la cual éste interaccionará con la tarea. Por ello, no es de sorprender que Petersen (2009) haya encontrado estudiantes que no solamente no respondieron a la consigna planteada, sino que además retaron abiertamente la temática propuesta en ésta.

4.2.2.4.3. LA PLANEACIÓN DE LA TAREA

En el caso de las tareas discretas, otro cuestionamiento que ha surgido al redactar las consignas se debe a si éstas deberían contener ideas o ejemplos a desarrollar. No obstante, como se mencionó previamente, entre más información esté incrustada en la consigna, mayor tendencia habrá para copiar información literalmente de ésta, por una parte, mientras que por

otra, la ejemplificación limita la capacidad creativa del sustentante. El estudio de Ong (2013) consistió en determinar si existe mayor riqueza de ideas cuando el examinando planea la tarea que cuando éstas cuentan con un tópico versus tópico + ideas. La autora identificó que los estudiantes explotan más su creatividad cuando sólo cuentan con el tópico y cuando no planean la tarea. En contraste, notó que el proceso de planeación induce también a que la riqueza de ideas se quede en el papel, pero que no se desarrollan en el texto, aunque la estructura y desarrollo de éstas es más profuso que aquellas que carecen de planeación. Finalmente, agrega la autora, una planeación extensa, aunque reduce las demandas cognitivas de la memoria de trabajo al momento de ir escribiendo, limita el tiempo de respuesta de la tarea e impide el desarrollo apropiado de ideas. De hecho, Johnson, Mercado y Acevedo (2012) encontraron que la planeación de la tarea escrita fomentaba medianamente la fluidez al escribir y no aumentaba la complejidad léxica ni gramatical.

4.2.2.4.4. EL TIEMPO DE ESCRITURA

Los estudios que se han conducido a este respecto han arrojado resultados que dificultan esclarecer el panorama con relación al tiempo que se asigna para la redacción de un ensayo. Knoch y Elder (2010) mencionan numerosos estudios, de los cuales en algunos se ha encontrado evidencia, aunque modesta, en favor de mayor tiempo asignado para la escritura. Sin embargo, comentan Knoch y Elder (2010) que estos estudios enfatizaron los resultados holísticos y no analíticos de las pruebas, además de que los resultados, aun cuando eran mejores cuando las condiciones de tiempo eran mayores, también eran sistemáticos; es decir, que todos los estudiantes (altos, intermedios y bajos) presentaban un mejor desempeño.

Debido a que numerosos docentes han cuestionado el tiempo que se dispone para escribir un ensayo bajo la modalidad cronometrada, la cual normalmente dura una hora para dos tareas, con el fin de medir la calidad de un texto en relación con el tiempo que se dispone para su redacción, Knoch y Elder (2010) condujeron un estudio en donde les asignaron dos tareas a los sustentantes, una que tenían que responder en 30 minutos y otra en 55 minutos. Los resultados demuestran que pese a que el contenido era menos elaborado en los ensayos de 30 minutos, los resultados obtenidos no eran significativamente diferentes en ambas pruebas. De hecho, destaca el hallazgo que en los ensayos de 30 minutos la cohesión y la complejidad gramatical fueron incluso mejores. Los autores infieren que probablemente esto se deba a que la presión para redactar en menor tiempo obliga a que los candidatos se enfoquen más en el tema que deben desarrollar y divaguen menos o pierdan el foco de los argumentos propuestos.

4.2.2.4.5. EL USO DE DICCIONARIO

Finalmente, otra de las discusiones se ha centrado en los beneficios del uso de un diccionario bilingüe o monolingüe para la resolución de las tareas escritas. Weigle (2002) comenta que el uso de diccionarios en las pruebas de lengua cronometradas no se permite debido a que el vocabulario es parte de lo que se pretende evaluar. La autora agrega que el uso de diccionarios tampoco es una habilidad que todos los sustentantes puedan desplegar satisfactoriamente, ya que dicha habilidad se restringe mayoritariamente a quienes son considerados como “buenos escritores”. En general, en la mayoría de los ECLAA no se permite el uso de éstos; sin embargo, la versión alemana del examen académico, *Test für Deutsch als Fremdsprache* (TestDaF), sí permite su uso. El estudio que East (2006) condujo con respecto al uso del diccionario partió del supuesto que el uso de un diccionario correspondería a una muestra más auténtica en los exámenes cronometrados, puesto que una de las críticas más asiduas a este tipo de exámenes se debe a que la falta de recursos genera una situación “artificial” de escritura que no se equipara con lo que un estudiante redacta en un contexto universitario.

The artificiality or lack of authenticity of timed writing tests lies essentially in their inability to reflect how writing is actually carried out in the real world. In particular, L2 writers in non-test situations have access to both extended time and resources (for example, dictionaries) to complete the task. Timed tests may facilitate “good measurement” but that measurement is not necessarily of “good writing” (East, 2006, p. 180).

De los hallazgos más sobresalientes del estudio de East (2006) destaca que el uso del diccionario incrementó considerablemente la sofisticación del léxico en los participantes. Sin embargo, una segunda mirada hacia la precisión en el uso del vocabulario demostró aproximadamente un 50% de errores en la selección de la palabra adecuada al contexto en el caso de los sustentantes con menores habilidades de escritura. Por lo tanto, si bien es cierto que el léxico fue más sofisticado, también es cierto que la mitad de éste fue seleccionado inapropiadamente, por lo cual su uso no aportaba ningún valor agregado a la escritura. En contraparte, los estudiantes con mayores habilidades de escritura sí se vieron beneficiados por el uso del diccionario, ya que las palabras que seleccionaron fueron apropiadas para el contexto. El autor concluye, siguiendo la línea de Weigle (2002), que el uso de diccionarios en los ECLAA conlleva un impacto negativo que incluso sobrepasa el beneficio de que los examinandos escriban con un léxico más sofisticado, pero poco pertinente.

4.2.2.4.6. EL MODO DE ESCRITURA

Una de las interrogantes que se presenta hoy en día al evaluar los exámenes cronometrados reside en por qué se hace a mano cuando los estudiantes escriben en computadora. Weigle (2002) e East (2009) comentan que los correctores tienden a evaluar con mayor dureza los textos redactados en computadora por varias razones: la primera, porque los textos se ven más cortos que los escritos a mano, segundo, porque no es posible evidenciar la competencia estratégica de la edición del texto (como podría ser cambiar una oración de lugar o borrar una palabra para buscar una más apropiada), y en tercer lugar, porque estamos acostumbrados a visualizar los textos escritos en computadora como productos finales, mientras que los redactados a mano son frecuentemente concebidos como borradores. Horakay, Bennet, Allen, Kaplan y Yan (2006) también identificaron mejores resultados, aunque no estadísticamente significativos, en los textos a mano, aun cuando los autores aducen esta diferencia a la familiaridad que los estudiantes presentan para escribir en computadora.

Mogey y Hartley (2013) reconocen estas mismas disimilitudes al momento de evaluar trabajos escritos a mano o en computadora, pero resaltan, por un lado, la importancia de generar conciencia en los estudiantes y correctores de qué tan largo es un ensayo escrito en computadora en comparación con aquellos escritos a mano. Por otro lado, destacan la necesidad de que los correctores se familiaricen con los textos escritos en computadora, vistos no como un trabajo final, sino más bien como un borrador que puede presentar errores. Más aún, los autores aseveran que en un futuro cercano, si no ya existente, los estudiantes redactarán los ensayos en computadora, incluso en los exámenes, y que los ensayos escritos a mano desaparecerán (Mogey y Hartley, 2013, p. 91).

Contradictoriamente al paulatino proceso de desaparición de la escritura a mano, en el estudio que Mogey, Paterson, Burk y Purcell (2010) condujeron para ver qué modo de escritura escogería un estudiante si se le diera la oportunidad de hacerlo, los autores identificaron que una gran mayoría optó por la redacción a mano (solamente 16 de 204 candidatos prefirieron hacerlo en computadora). Al parecer algunas de las posibles causas en elegir la escritura a mano se deben a la posible dificultad de escribir en computadora, a la desconfianza con la tecnología y también a que se trataba de estudiantes que recién habían ingresado a la universidad, quienes pudieron haber visualizado la escritura en computadora como “demasiado riesgosa”. Pese a la contundente preferencia en el modo de escritura a mano, los autores concluyen que la posibilidad de elegir el modo de escritura debería quedar al arbitrio de los examinandos. Lo cierto es que en

realidad los estudiantes escriben a mano y en computadora durante sus estudios universitarios, tanto notas de clase como exámenes o ensayos, razón por la cual valdría la pena considerar la posibilidad de que ellos elijan el modo de escritura que mejor les acomode. Aunado a lo anterior, el uso del teclado en lengua extranjera causa dificultades debido a que incluso con el uso del mismo alfabeto, la ubicación de los caracteres en un teclado en español, francés, alemán o inglés no son iguales; por lo tanto, la velocidad de escritura al usarlo dependerá de la familiarización que el sustentante tenga con éste.

4.3. MÉTODOS E INSTRUMENTOS

La primera inferencia planteada en esta investigación se refiere a la definición del ámbito y a la pertinencia de las tareas de escritura incluidas en la PE del EXLEAA.

- **Las tareas de la PE incluidas en el EXLEAA son pertinentes para evaluar las habilidades de escritura académica.**

Para ello, lo primero fue conocer el ámbito para el cual deseamos evaluar. Sin embargo, debido a que el campo de la escritura académica es vasto, se condujo un estudio exploratorio para conocer solamente dos aspectos del ámbito académico: la importancia de la escritura a nivel posgrado y las habilidades de escritura que, según docentes del posgrado, requieren sus estudiantes para ser académicamente competentes. Por lo menos en el contexto mexicano, las prácticas de evaluación docente en el ámbito universitario han sido poco estudiadas. Más aún, las tipologías textuales que les solicitan los docentes de posgrado a sus estudiantes representan aún un terreno más árido en el contexto hispanohablante. De ahí que, se optara por un estudio exploratorio de tipo descriptivo. Los estudios exploratorios resultan particularmente útiles cuando se trata de “examinar un tema o problema de investigación poco estudiado, del cual se tienen muchas dudas o no se han abordado antes” (Hernández, Fernández, Baptista, 2010, p. 152).

Posteriormente, con base en los resultados obtenidos en el estudio exploratorio y la revisión bibliográfica de este apartado respecto de las tareas de escritura de los ECLAA, se llevó a cabo una discusión teórica de la pertinencia de las tareas. En campos tan amplios, como el académico, es importante que las tareas diseñadas sean representativas del universo para el cual se pretende evaluar. La PE fue diseñada partiendo de los requerimientos de escritura de los

estudiantes universitarios; no obstante, las prácticas de evaluación varían considerablemente de una disciplina a otra y también de un grado a otro.

4.3.1. LOS PARTICIPANTES

Con información proveniente de la página del posgrado de la UNAM y de cada uno de los distintos posgrados e institutos de investigación, se generó una base de datos con los correos electrónicos de 800 docentes pertenecientes a los 40 programas de posgrado con que cuenta la UNAM. De forma equitativa y totalmente aleatoria, se seleccionó a veinte docentes de cada uno de los programas de posgrado. Con el fin de captar a la mayor cantidad de participantes, la invitación a contestar el cuestionario se envió por correo electrónico el 16 de enero de 2014, durante el período inter semestral. Una semana más tarde, se les envió un recordatorio de participación. De los 800 docentes de posgrado invitados a participar en el estudio, se obtuvo respuesta de 63, pertenecientes a 27 programas de posgrado. El Cuadro 4.3 muestra un resumen de las áreas participantes y los posgrados a los que pertenecían los docentes. En el Anexo 1.1 figura la lista completa de los docentes que participaron, así como su procedencia y la materia que impartían al momento de recibir la encuesta.

Cuadro 4.3: Posgrados participantes

Ciencias Físico-Matemáticas e Ingenierías	Ciencias Biológicas, Químicas y Salud
Ingeniería en sistemas computacionales Ingeniería industrial Ciencias Físicas Ciencias de la tierra Ciencia e Ingeniería de Materiales Astrofísica	Neurobiología Ciencias biomédicas Ciencias Médicas, Odontológicas y de la Salud Facultad de Psicología Ciencia Bioquímica Ciencias Biológicas Ciencias del Mar y Limnología
Ciencias Sociales	Humanidades y Artes
Ciencias de la Administración Ciencias Políticas y Sociales Geografía Derecho Antropología	Letras Estudios Mesoamericanos Historia del arte Arte y Diseño Arquitectura Lingüística Aplicada Música Bibliotecología y Estudios de la Información Posgrado en Pedagogía

4.3.2. EL INSTRUMENTO

En *google forms*, se diseñó un cuestionario abierto sobre las prácticas que los docentes emplean para evaluar el desempeño académico de sus estudiantes de posgrado (ver Anexo 1.2) y una encuesta en escala tipo Likert sobre las habilidades de escritura requieren los estudiantes de posgrado para ser académicamente competentes (ver Anexo 1.3).

Cabe mencionar que la información del cuestionario era opcional, por lo cual el docente bien podía dejar alguna respuesta en blanco. El objetivo primordial de éste fue identificar tres aspectos: la importancia y el peso de la escritura en la evaluación, las tipologías textuales solicitadas (en caso de evaluar mediante trabajos escritos), y los criterios que los docentes emplean para evaluar los trabajos escritos de sus estudiantes. Para poder identificar las habilidades de escritura importantes para el éxito académico, se replicó el estudio que Rosenfeld et al. (2004) llevaron a cabo con miras a la modificación del TOEFL PBT por su nueva versión iBT, implementada en 2006.

4.3.3. EL PROCEDIMIENTO

Tras un primer pilotaje con cinco docentes, se afinaron las preguntas planteadas en el cuestionario y se redujo la lista de 39 descriptores (originalmente incluidos en el estudio de Rosenfeld et al. (2004)) a 33. La razón por la cual se descartaron seis descriptores se debió a dos razones: o éstos no representaban habilidades de escritura per se (utilizar el software de procesamiento de textos para planificar y crear textos, trabajar de forma independiente para planificar y redactar textos y trabajar en colaboración para planificar y redactar textos) o presentaban habilidades que puntuaron demasiado bajo –al igual que en el estudio de Rosenfeld et al.– quizás por tratarse de habilidades de escritura características de unos cuantos programas académicos (analizar los significados en una pieza de literatura imaginativa, escribir persuasivamente apelando principalmente a las emociones del lector, experiencias y valores éticos, y describir y evaluar la eficacia de las estrategias técnicas-retóricas de un escritor). A diferencia del cuestionario, en el caso de la encuesta, la respuesta de cada uno de los descriptores era obligatoria; de lo contrario, el sistema no permitía que se enviara la información. Una vez que se afinaron el cuestionario y la encuesta, se les envió a los docentes del posgrado.

4.3.4. EL ANÁLISIS

La información del cuestionario fue sistematizada en EXCEL mediante porcentajes para ver qué peso tiene la escritura en cada una de las cuatro áreas del saber. También se extrajo

información importante proporcionada por los docentes sobre aspectos relacionados con la escritura académica. En el caso de la encuesta, como se trataba de un estudio de tipo exploratorio, solamente se llevaron a cabo estudios descriptivos, en SPSS versión 21, que permitieran dar un primer acercamiento a cuáles son las prácticas de evaluación docente a nivel posgrado. Se obtuvieron las medias de las respuestas de los docentes con el fin de poder determinar cuáles son las habilidades de escritura que éstos consideran más importantes para el éxito académico de sus estudiantes. Con la información procedente de la fase descriptiva, se pudo evidenciar si las tareas de la PE eran pertinentes para el ámbito académico. Kane (2002b, 2006, 2013) y Kane et al. (1999) comentan que lo ideal en los exámenes de lengua de alto impacto será lograr un balance entre la autenticidad de la tarea y también entre la estandarización de los exámenes que permitan generalizar para el ámbito para el cual se desea evaluar.

4.4. RESULTADOS

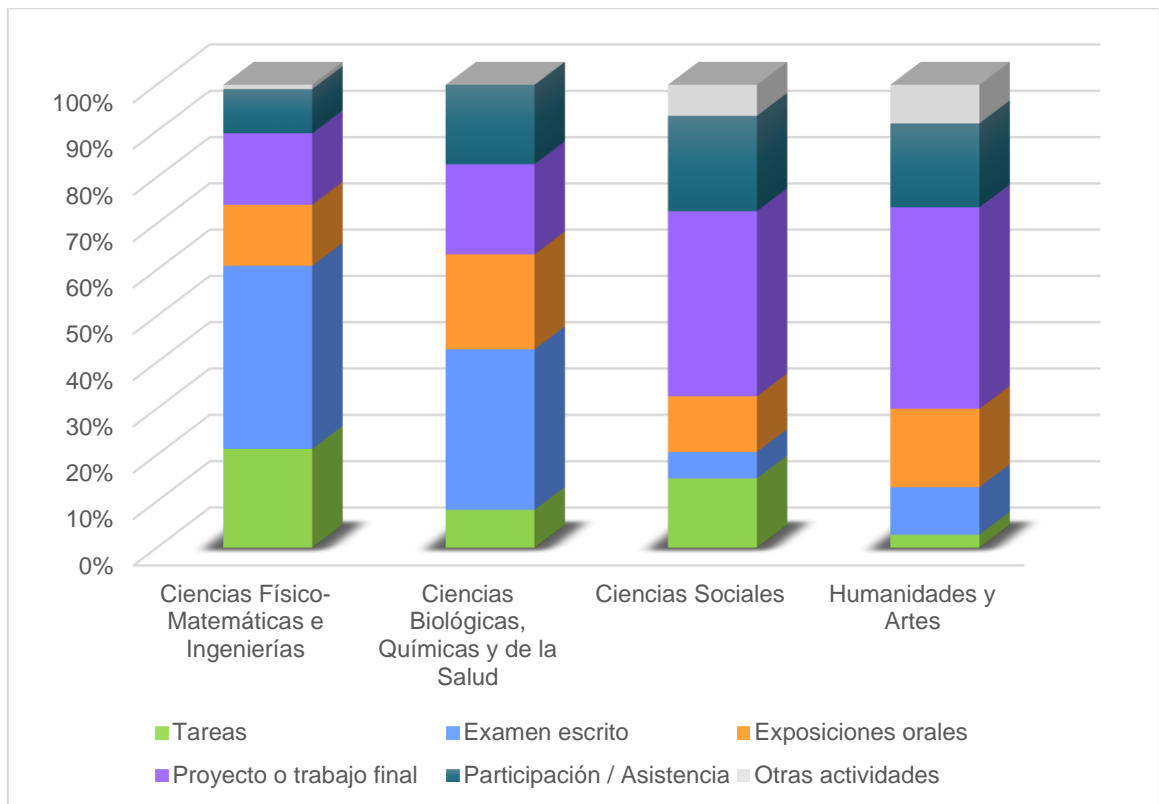
En esta sección se reportan los resultados del estudio que se condujo con los docentes de posgrado que participaron en el estudio sobre la importancia de la escritura académica y de las habilidades de escritura dentro del contexto de posgrado de sus estudiantes. De los 62 docentes que respondieron el cuestionario, los datos de uno de ellos fueron eliminados completamente de la base de datos porque no se contaba con información del posgrado de procedencia ni de la materia impartida. Con respecto a la encuesta tipo Likert, se eliminó a cinco participantes de ésta, pues respondieron que todas las habilidades de escritura eran “extremadamente importantes”; dato que, aunque incrementaba ligeramente la importancia de cada una de las habilidades de escritura, no aportaban información discriminadora entre las habilidades ni generaba ninguna diferencia en la jerarquización de éstas.

Los resultados aquí mostrados son de carácter exploratorio y descriptivo; es decir, pretenden dar una primera aproximación al estudio de las habilidades de escritura a nivel posgrado. La muestra recolectada es pequeña en relación con la cantidad de docentes que imparten cursos en el posgrado de la UNAM; no obstante, la amplia gama de posgrados de procedencia, así como la variedad de materias impartidas, permiten este primer acercamiento y familiarización con un tema poco estudiado y poco conocido en la UNAM y en las universidades hispanoamericanas. No obstante, la jerarquización de las habilidades de escritura consideradas como primordiales para los estudiantes de maestría hacen eco a los estudios conducidos por Rosenfeld et al. (2004).

4.4.1. EL CUESTIONARIO SOBRE PRÁCTICAS DE EVALUACIÓN DOCENTE

En la Gráfica 4.2 se muestran las cuatro áreas del saber en la que se encuentran los 40 posgrados de la UNAM. Primeramente, se describen los datos más sobresalientes y más adelante se hace una relación detallada de lo más significativo de cada una de ellas. Como se puede apreciar en la Gráfica 4.2, los exámenes escritos representan la práctica de evaluación frecuentemente empleada por los docentes del área 1 (Ciencias Físico-Matemáticas e Ingenierías). En el caso del área 2 (Ciencias Biológicas, Químicas y de la Salud), además de los exámenes escritos, son igualmente relevantes los exámenes/exposiciones orales. Sin embargo, en el caso del área 3 (Ciencias Sociales) y área 4 (Humanidades y Artes), los proyectos o trabajos finales –generalmente a modo de ensayo– figuran como la forma de evaluación más representativa.

Gráfica 4.2: Prácticas de evaluación docente en el Posgrado de la UNAM



Ahora bien, con respecto a las demás modalidades de evaluación, en el Área 1, de manera equitativa, los exámenes orales, las tareas y los trabajos finales figuran en segunda posición; mientras que en el Área 2, los trabajos finales y la participación se ubican en segundo término

de importancia, desplazando a las tareas en tercera posición. En cuanto a las Áreas 3 y 4, figuran en segundo término la participación y la asistencia. En ambas áreas, las demás prácticas de evaluación quedan relegadas a tercer término. A continuación, se detallan los rasgos que más destacan con respecto a los trabajos escritos que los docentes del posgrado solicitan para evaluar el conocimiento de sus estudiantes. En el Anexo 1.4 se encuentra una relación detallada de cómo contestaron los docentes de los diferentes posgrados.

4.4.1.1. RESULTADOS DEL ÁREA 1: CIENCIAS FÍSICO-MATEMÁTICAS E INGENIERÍAS

De acuerdo con las respuestas proporcionadas por los docentes de esta área, tanto en los exámenes como en las tareas es frecuente la descripción de conceptos (entre media cuartilla y dos) y problemas matemáticos a desarrollar. En algunos casos, la resolución de problemas matemáticos incluye la interpretación o explicación de los mismos. En el caso de lecturas, se piden ensayos con críticas o comentarios sobre las mismas o bien se elaboran preguntas abiertas para que desarrollen conceptos. Asimismo, se solicita la elaboración e interpretación o explicación de gráficas. En el caso de los proyectos y el estudio de caso reportado, no se describe puntualmente qué tipo de información se incluye en éstos, salvo la incorporación de gráficas con su respectiva interpretación. Finalmente, con respecto a los criterios para evaluar, los docentes ponen especial énfasis en la capacidad de reflexión y en la relevancia del contenido. Asimismo, añaden otros criterios como: la correcta aplicación de conceptos, el desarrollo de nuevas ideas y la exactitud matemática.

Ejemplo de cómo contestó uno de los docentes de esta área:

Evaluación: 50% Tareas (3) y 50% Examen

Tarea 1: Cuatro problemas que necesitan algunos cálculos analíticos y aritméticos, y por lo menos un párrafo de interpretación. Una pregunta para explicar tres conceptos clave.

Tarea 2: Dos problemas que requieren muy pocos cálculos, y la elaboración de programas para realizar los cálculos necesarios y gráficas. Un cuestionario de ocho preguntas sobre una lectura.

Tarea 3: Cinco problemas que requieren tanto cálculos como interpretaciones.

Examen: Se plantearon 8 preguntas (1.25 pts.) y 4 problemas (2.5 pts.). El estudiante podía elegir a qué responder sabiendo que el examen sería calificado sobre 10 puntos y con la obligación de responder por lo menos a dos preguntas y a un problema. Las respuestas a las preguntas necesitaban de aproximadamente media cuartilla. Para los problemas, depende del método elegido y de la cantidad de detalle que el estudiante incluya.

4.4.1.2. RESULTADOS DEL ÁREA 2: CIENCIAS BIOLÓGICAS, QUÍMICAS Y DE LA SALUD

En lo que respecta a las prácticas de escritura, los trabajos no hacen énfasis en el número de cuartillas, sino en que éstos sean concretos respecto del tema a desarrollar. En general, éstos se enfocan a la redacción de parte de la tesis, artículos de investigación y proyectos. Los exámenes pueden ser de opción múltiple y/o con preguntas a desarrollar. En casos puntuales, los profesores reportan haber solicitado una reseña de 1 ó 2 cuartillas y un ensayo relacionado con el tema de tesis, de aproximadamente 10 cuartillas. En realidad, en el área dos destacan las evaluaciones a través de presentaciones orales. Finalmente, dentro de los criterios para evaluar, se enfatiza la relevancia del contenido, el uso apropiado de las fuentes, la capacidad de reflexión y la coherencia y cohesión, pasando a segundo término la precisión gramatical y el formato de presentación.

Ejemplo de cómo contestó uno de los docentes de esta área:

Evaluación: Examen 10 %, Asistencia y participación en clase 10 %, Presentación oral en clase 50%, Tareas 30 %

Siempre preparo una tarea (o examen) con 80 % de formato selección (seleccionar una respuesta correcta de varias posibles respuestas) y 20 % de escribir la respuesta. En caso de pedir un escrito, no pido ninguna restricción de número de palabras. En general, pido a los alumnos que escriban de manera muy breve y concreta.

4.4.1.3. RESULTADOS DEL ÁREA 3: CIENCIAS SOCIALES

En cuanto a la escritura, destacan los proyectos de investigación y los trabajos finales con extensiones entre 10 y 50 cuartillas. Los trabajos pueden ser avance de tesis, proyectos de investigación, ensayos argumentativos, trabajos críticos, análisis de artículos de investigación con una extensión aproximada entre 10 y 15 cuartillas. En relación con los criterios que emplean para evaluar, destacan la relevancia del contenido, el uso apropiado de las fuentes, la capacidad de reflexión, la coherencia y la cohesión. En menor medida consideran la precisión gramatical, la presentación y formato de acuerdo al género discursivo.

Ejemplo de cómo contestó uno de los docentes de esta área:

Evaluación: Asistencia y participación 30%, Presentaciones orales 30%, Ensayo final 40%

El ensayo final consta de un ensayo crítico sobre la lectura de un libro previamente asignado. La extensión es de un mínimo de 5 cuartillas y máximo de 15, escrito en letra times a 1.5 espacio.

4.4.1.4. RESULTADOS DEL ÁREA 4: HUMANIDADES Y ARTES

En relación con los trabajos escritos, destacan los ensayos argumentativos entre 12 y 20 cuartillas en los cuales se valora ampliamente el análisis crítico y reflexivo de los estudiantes respecto de un cierto tema o lectura asignada. En el caso de secciones o capítulos de tesis, la extensión varía dependiendo del contenido a desarrollar, aunque también destaca la perspectiva argumentativa dentro de éstos. En otras ocasiones, se solicita el desarrollo de un proyecto de investigación con la metodología apropiada. Los trabajos cortos no figuran como primordiales dentro de la evaluación, sin embargo, se hace referencia a cuestionarios abiertos, reseñas y otros trabajos de menos de cinco cuartillas. En relación con los criterios para evaluar trabajos escritos, además de los incluidos en la lista, destaca el análisis y la reflexión crítica, así como la creatividad intelectual, lo cual está en sintonía con respecto a la naturaleza de los trabajos solicitados.

Ejemplo de cómo contestó uno de los docentes de esta área:

Evaluación: Asistencia y participación en clase: 25%, Exposición individual: 25%, Examen oral final: 25%, Trabajo final: 25%

El trabajo consistió en un capítulo de la tesis de maestría o doctorado que cada alumno esté realizando. A veces, por hallarse en los comienzos de su investigación, dicho capítulo es en realidad el proyecto explícito del trabajo que habrán de realizar.

4.4.2. ENCUESTA SOBRE LA IMPORTANCIA DE LAS HABILIDADES DE ESCRITURA ACADÉMICA

Con respecto a la encuesta tipo Likert, la muestra quedó conformada por 57 docentes de la siguiente manera: Ciencias Físico-Matemáticas e Ingenierías, 10 docentes; Ciencias Biológicas, Químicas y Salud, 12 docentes; Ciencias Sociales, 11 docentes; y Humanidades y Artes, 24 docentes. Ahora bien, el análisis de una escala tipo Likert ha sido controvertida, pues por una parte hay quienes tratan los datos como si fueran de tipo escalar o de intervalo (Creswell, 2005, citado en Hernández et al., 2010, p.251), puesto que éstos han sido probados en múltiples ocasiones (tal fue el caso del estudio llevado a cabo por Rosenfeld et al., 2004); sin embargo, en sentido estricto, autores como Jamieson (2004, citado en Hernández et al., 2010, p.251), mencionan que los datos de una escala tipo Likert son en realidad jerárquicos u ordinales; razón por la cual no es posible establecer las medias, sino más bien se determina la medida de tendencia central (moda) o la sumatoria de cada una de las frases de Likert para después jerarquizarlas. En el presente trabajo, debido a que uno de los objetivos consistió en comparar las habilidades de escritura entre los docentes de la UNAM con su contraparte estadounidense, se analizó la encuesta mediante la obtención de sus medias.

El Cuadro 4.4 presenta en orden descendente aquellas habilidades de escritura que los docentes del posgrado consideraron como más importantes para el correcto desempeño académico de los estudiantes. Como se puede apreciar en este cuadro, se jerarquizaron de mayor a menor las habilidades identificadas dentro del rango de muy importantes a extremadamente importantes (de 4.0 a 5.0) y del rango de importantes a muy importantes (de 3.0 a 3.99). No hubo habilidades consideradas en un rango inferior, quizás porque previamente habían sido eliminadas durante el pilotaje, aunque también es cierto que en este tipo de encuestas es raro que los docentes consideren que ciertas habilidades no son importantes para los estudiantes.

Cuadro 4.4: Jerarquización de las habilidades de escritura por parte de los docentes del posgrado de la UNAM

Habilidades de escritura	Media
20. Dar crédito adecuado a las fuentes.	4.77
23. Organizar las ideas y la información coherentemente.	4.72
33. Evitar errores ortográficos.	4.72
24. Escribir de forma precisa y concisa, evitando frases vagas o vacías.	4.70
3. Abstractar o resumir información esencial.	4.70
16. Presentar los datos y la información de manera clara y lógica.	4.67
6. Analizar y sintetizar información de múltiples fuentes (incluye comparación y contraste).	4.63
5. Explicar un hecho o suceso utilizando evidencia.	4.58
29. Revisar y editar el texto para mejorar su calidad, coherencia y precisión.	4.54
19. Integrar el material citado y referenciado debidamente dentro del texto.	4.53
26. Escribir con fluidez, evitando que el lenguaje sea complicado y rebuscado.	4.53
32. Utilizar correctamente la gramática del español escrito con el fin de evitar que el lector se distraiga o que se cometan errores que alteren el significado.	4.51
12. Interpretar los datos dentro de un marco determinado.	4.47
8. Escribir persuasivamente mediante la construcción de un argumento bien razonado para apoyar o refutar una postura.	4.46
30. Usar vocabulario técnico y específico al contenido del tema con precisión y de forma apropiada para un propósito y audiencia en particular.	4.40
10. Examinar el razonamiento en un argumento dado y discutir las fortalezas y debilidades de su lógica.	4.33
9. Explorar las relaciones entre ideas complejas y posiblemente contradictorias.	4.32
31. Elegir palabras efectivamente.	4.28
2. Explicar cómo se lleva a cabo un procedimiento.	4.19
15. Escribir adecuadamente para un público bien informado y reflexivo.	4.16
21. Desarrollar un debate bien enfocado y debidamente sustentado mediante ejemplos relevantes.	4.16
11. Identificar los problemas en una línea de acción propuesta o en la interpretación de hechos y proponer soluciones o alternativas de interpretación.	4.12
1. Describir observaciones.	4.07

25. Escribir con claridad mediante transiciones suaves de un pensamiento a otro.	4.05
4. Expresar opiniones personales.	4.05
22. Aclarar las relaciones entre ideas principales y secundarias.	4.05
13. Clasificar la información de acuerdo con categorías o jerarquías.	3.88
7. Predecir consecuencias o resultados mediante el análisis de información, patrones o procesos.	3.88
27. Variar la estructura de las oraciones para comunicar ideas de manera efectiva.	3.81
18. Utilizar formatos eficientes y claros para organizar la información y guiar al lector.	3.75
14. Utilizar las convenciones de un género en particular.	3.74
28. Expresar ideas en forma original o novedosa para mantener el interés del lector.	3.70
17. Usar la analogía, metáfora o comparación para definir o explicar conceptos técnicos o abstractos para una audiencia general.	3.26

Contrastando los resultados obtenidos con los de Rosenfeld et al. (2004, pp.14 y 15), se puede apreciar que los docentes de la UNAM evaluaron las habilidades de escritura con puntajes ligeramente más elevados. Sin embargo, la manera de jerarquizarlas es muy semejante. El Cuadro 4.5 muestra el comparativo de las habilidades con mayor importancia entre los docentes de posgrado de la UNAM y de los docentes estadounidenses encuestados por Rosenfeld et al. (2004). Como se evidencia en este cuadro, de las once habilidades más importantes, la coincidencia de importancia en ambos grupos es de ocho y con una jerarquía muy semejante.

Finalmente, cabe resaltar los valiosos comentarios de los docentes con respecto a la importancia de la escritura académica en el posgrado (ver Anexo 1.5), quienes además enfatizan la instrucción de la escritura académica a toda la comunidad universitaria, sobre todo a los estudiantes mexicanos. Asimismo, señalan que la necesidad de escribir en inglés en numerosos posgrados está acrecentando las dificultades de escritura que encaran los estudiantes de posgrado; razón por la cual tienen a bien en apuntar la importancia en la impartición de cursos de escritura académica dentro de las mismas facultades.

Cuadro 4.5: Comparativo entre las habilidades de escritura más importantes para los docentes de posgrado de la UNAM y de EEUU

Docentes de la UNAM	Media	Docentes de EEUU	Media
Dar crédito adecuado a las fuentes.	4.77	Dar crédito adecuado a las fuentes.	4.5
Organizar las ideas y la información coherentemente.	4.72	Organizar las ideas y la información coherentemente.	4.4
Evitar errores ortográficos.	4.72	Utilizar correctamente la gramática del inglés escrito con el fin de evitar que el lector se distraiga o que se cometan errores que alteren el significado.	4.4
Escribir de forma precisa y concisa, evitando frases vagas o vacías.	4.70	Evitar errores ortográficos.	4.3
Abstraer o resumir información esencial.	4.70	Abstraer o resumir información esencial.	4.2
Presentar los datos y la información de manera clara y lógica.	4.67	Analizar y sintetizar información de múltiples fuentes (incluye comparación y contraste).	4.2
Analizar y sintetizar información de múltiples fuentes (incluye comparación y contraste).	4.63	Integrar el material citado y referenciado debidamente dentro del texto.	4.2
Explicar un hecho o suceso utilizando evidencia.	4.58	Desarrollar un debate bien enfocado y debidamente sustentado mediante ejemplos relevantes.	4.2
Revisar y editar el texto para mejorar su calidad, coherencia y precisión.	4.54	Escribir con claridad mediante transiciones suaves de un pensamiento a otro.	4.2
Integrar el material citado y referenciado debidamente dentro del texto.	4.53	Escribir de forma precisa y concisa, evitando frases vagas o vacías.	4.2
Escribir con fluidez, evitando que el lenguaje sea complicado y rebuscado.	4.53	Revisar y editar el texto para mejorar su calidad, coherencia y precisión.	4.2

4.5. DISCUSIONES, LIMITACIONES Y FUTURAS INVESTIGACIONES

En este apartado se abordan en primer lugar las discusiones sobre la importancia de la escritura académica en el posgrado y después la pertinencia de las tareas de la PE del EXELEAA. Posteriormente se describen las limitaciones y se proponen futuras investigaciones.

4.5.1. LA DESCRIPCIÓN DEL ÁMBITO: LA ESCRITURA DENTRO DEL CONTEXTO ACADÉMICO

La escritura académica juega un papel central dentro del contexto universitario tanto para la evaluación del conocimiento de los estudiantes como para la difusión del mismo a través de tesis y de artículos de investigación. Lo anterior ha posicionado a la escritura como la habilidad lingüística que requiere mayor atención y que requiere reforzamiento por parte de las diversas facultades; ya sea mediante cursos generales o como lo propone Carlino (2004a, 2005b, 2007, 2013), integrada a cada uno de los planes curriculares de las facultades.

En cuanto a las prácticas de evaluación docentes, se evidencia claramente el valor de la escritura en cualesquiera de las cuatro áreas del saber y a través de los diferentes posgrados que se ofertan en la UNAM. En el área Físico-Matemáticas y de las Ingenierías, destaca el uso de la escritura académica como herramienta utilizada en exámenes, en proyectos y también en la redacción de la tesis. Dicha habilidad, de forma muy generalizada, coadyuva a la descripción de gráficas y también de conceptos matemáticos y/o numéricos que por sí mismos no explicarían determinados fenómenos naturales. En contraste, en el área de Ciencias Sociales, Humanidades y Artes, dicha habilidad se enfoca primordialmente en la elaboración de ensayos, trabajos críticos y argumentativos que pretenden manifestar diferentes posturas y establecer puntos a favor o en contra de éstas.

De forma peculiar, la asistencia y participación de los estudiantes, aún a nivel posgrado, juega un papel primordial en la evaluación del desempeño académico. Esto da cuenta de la importancia que conlleva la evaluación continua del conocimiento dentro del salón de clases. En contraste, la habilidad oral como forma de evaluación a través de exámenes o exposiciones queda relegada en tercer término excepto en el área de Ciencias Biológicas, Químicas y de la Salud; no obstante, dicha habilidad resulta esencial tanto dentro de la evaluación continua como en la de aprovechamiento. Cabe añadir que el uso de tecnologías de la información como foros de discusión u otros mecanismos que pudieran emplearse paralelamente para evaluar el conocimiento, son prácticamente nulos. Sería necesario indagar a profundidad si los docentes emplean alguna herramienta tecnológica para fomentar discusiones o incluso para evaluar a sus estudiantes, aunque grosso modo parece no ser una práctica recurrente en el posgrado de la UNAM.

En cuanto a las habilidades de escritura académica que requiere un estudiante para ser académicamente competente, destaca en primer lugar la estrecha relación entre los resultados obtenidos por parte de los docentes de la UNAM y aquellos observados por Castelló et al. (2012) en el contexto hispanohablante y por los de Rosenfeld et al. (2004) en el contexto anglófono. Lo anterior permite ver que dichas habilidades van más allá de los requerimientos de la UNAM o de cualquier otra institución de nivel superior. De manera puntual, vale la pena resaltar la habilidad que figura en primer término, pues a nivel posgrado resulta incuestionable la acreditación adecuada de las fuentes empleadas para la construcción del conocimiento. Por ende, más allá de la mera línea discursiva que gira en torno al plagio académico, enseñar dicha habilidad resulta crucial para que los estudiantes universitarios y, sobre todo los de posgrado, puedan generar conocimiento.

Angulo (2013) hace mención detallada de los problemas por los cuales atraviesa el estudiante universitario al momento de dar crédito adecuado a las fuentes y de la falta de voz y de presencia en su discurso debido a que generalmente el estudiante se limita a citar autoridades sin el entramado intertextual que dé cuenta de su postura ante cierto fenómeno. En este sentido, valdría la pena reflexionar en torno a dos preguntas: ¿de quién es la responsabilidad de enseñar el proceso de la construcción del conocimiento a nivel posgrado?, y ¿es justo exigirle a un estudiante de posgrado que sea competente en el dominio de las habilidades de escritura –como para publicar artículos científicos o académicos– cuando no se le han enseñado explícitamente dichas habilidades? Por esta razón, sería importante desarrollar cursos de escritura académica en cualesquiera de los ciclos universitarios.

Las once habilidades consideradas como primordiales parecen abordar dos líneas: la competencia comunicativa escrita (organizar las ideas y la información de forma coherente, evitar errores ortográficos, escribir de forma precisa, presentar los datos de forma clara, revisar y editar el texto, escribir con fluidez y utilizar correctamente la gramática del español) y la construcción del conocimiento (dar crédito a las fuentes, abstraer información esencial, analizar y sintetizar información de múltiples fuentes, e integrar el material citado y referenciado debidamente dentro del texto). Así pues, podemos ver cómo la competencia comunicativa escrita se encuentra íntimamente relacionada con la capacidad para construir el conocimiento. Por esto, como mencionan Carlino (2004a, 2005b, 2007, 2013), Flores (2014) y Zhizhko (2014), enseñar cursos de redacción desarticulados de las necesidades académicas de cada una de las facultades resulta fútil. Entonces, La incorporación de los cursos académicos deberían abordar estos dos

ejes (la competencia comunicativa escrita y la construcción del conocimiento) insertos en las tipologías textuales que son características de las distintas comunidades discursivas del posgrado. Teberosky (2007, p.17) comenta que “escribir textos académicos es una actividad de construcción del texto propio, a partir de textos ajenos (intertextualidad), que da lugar a un producto final fruto del desarrollo de distintas posiciones enunciativas: escriba, escritor, compilador, lector de textos ajenos y propios”.

Por último, resaltan los valiosos comentarios vertidos por los docentes que manifiestan, por una parte, que el nivel que presentan los estudiantes que ingresan al posgrado es bajo, mientras que por otra, insisten en que los cursos de escritura académica son esenciales dentro del contexto universitario en sus diversos ciclos: licenciatura, maestría y doctorado. Por último, vale la pena resaltar los comentarios respecto de la necesidad de escribir en inglés, particularmente en las Ciencias Físico-Matemáticas e Ingenierías, en donde gran parte del conocimiento se publica en esta lengua. Por esta razón, la falta de herramientas para poder escribir adecuadamente en lengua materna acrecienta aún más las dificultades para hacerlo en otra lengua. En suma, el estudio exploratorio aquí descrito permite un primer acercamiento al complejo fenómeno de la escritura académica, de los retos que conlleva y de las habilidades de escritura que son primordiales dentro del posgrado universitario.

4.5.2. LA PERTINENCIA DE LAS TAREAS DE LA PE DEL EXELEAA

Primeramente, cabe mencionar que al seleccionarse tareas discretas, la posibilidad de evidenciar la construcción del conocimiento queda fuera. Entonces, resulta importante delimitar en primera instancia el alcance de la PE del EXELEAA, ya que como se evidenció en la descripción del ámbito, los docentes consideran que las habilidades de escritura primordiales para que los estudiantes universitarios sean académicamente competentes giran fundamentalmente en torno a dos grupos: las habilidades de escritura necesarias para construir conocimiento y las habilidades de escritura de índole académico para comunicar efectivamente dicho conocimiento. Sin embargo, valdría la pena considerar lo siguiente: ¿es posible evidenciar la construcción del conocimiento en un texto de aproximadamente 150 palabras con el input proporcionado por uno o dos textos breves? Y, en el contexto hispanohablante, ¿siempre construyen los estudiantes el conocimiento de fuentes provenientes del español? En todo caso, habría que buscar tareas que se asemejan más al uso de la lengua en el ámbito académico; es decir, a una escritura académica que refleje las habilidades de escritura que son importantes en el contexto académico (p.ej. tareas que involucran la lectoescritura de los estudiantes).

Con respecto a las gráficas, resulta evidente que la interpretación de éstas no siempre es sencilla. Sin embargo, debido al creciente uso de éstas en los artículos de investigación, incluso en las Ciencias Sociales, y Humanidades y Artes, resulta fundamental incluirlas, pues representan parte de la literacidad académica universitaria. Asimismo, la gran ventaja del estímulo visual radica en que éste contiene los elementos necesarios para que los estudiantes escriban, con lo cual se evita que copien fragmentos de los textos fuente en tareas integradas. No obstante, la selección de gráficas no es una tarea fácil, pues es importante contemplar que ésta desplegará cierto léxico que el examinado debe poseer, sin importar los estudios académicos que desee cursar o esté cursando. Por esta razón, se deberá incorporar información visual que no contenga léxico especializado (Yang, 2012), dado que esto limitará la capacidad del candidato para explotar al máximo el potencial de escritura y además coartará la capacidad de escritura en aquellos candidatos con menor aptitud. Entonces, resulta importante seleccionar gráficas de barra o de pastel que sean sencillas y claras para los sustentantes. Finalmente, lo ideal sería que quienes seleccionan las gráficas, así como los evaluadores, respondieran la tarea, puesto que solamente así podrían identificar posibles problemas que los candidatos pudieran tener al momento de escribir y, consecuentemente, tomar acciones previas a la aplicación de la prueba.

En el caso de los ensayos de carácter argumentativo, como se pudo apreciar en la revisión bibliográfica, el contenido de éstos permitirá que se establezca o no un lazo interactivo entre el sustentante y la tarea. Por ello, más que abogar a favor o en contra de este tipo de tareas escritas, resulta importante incorporar temas que resulten interesantes para los candidatos, quienes claramente poseen un sentido crítico y están dispuestos a participar siempre y cuando los temas estimulen su capacidad de análisis. Entonces, no resulta trivial que Petersen (2009) haya encontrado casos en los cuales los estudiantes se opusieron abiertamente a contestar este tipo de ensayos, dado que cuando los alumnos perciben que su capacidad crítica se ve minimizada con temas absurdos o fuera de lugar, éstos reaccionan negativamente.

En relación con el modo de escritura, es evidente que la escritura a mano distorsiona la realidad del texto y favorece o demerita la apreciación subjetiva por parte del evaluador. Sin embargo, la costumbre de evaluar textos escritos a mano dificulta el cambio a aquellos escritos en computadora, dado que parecen ser más cortos y productos finales revisados y editados, cuando en realidad no es así. Aunado a lo anterior, la estrategia de corregir en el papel ha sido siempre vista como un aspecto positivo de la escritura en segundas lenguas, misma que se

esfuma al no poder evidenciar lo que el sustentante ha editado en una computadora. Otra cuestión importante radica en que los participantes que escribieron sus textos en computadora ya son alumnos de posgrado y, por lo tanto, están familiarizados con el teclado en español; sin embargo, si un candidato que apenas desea ingresar al ciclo universitario no estuviera familiarizado con éste, posiblemente se vería en desventaja, dependiendo de la práctica que haya tenido previamente con el teclado en el idioma que desea escribir. Entonces, valdría la pena proponer una modalidad abierta; es decir, que el candidato escogiera la manera mediante la cual escribirá sus textos (véase Horkay, Bennett, Allen, Kaplan y Yan, 2006).

Retomando a Weigle (2002), pese a que las variables que se ven involucradas en las tareas de escritura no esclarecen el panorama de cuáles serían óptimas para evaluar en un contexto académico, éstas no son insuperables. De hecho, entre mayor sea el conocimiento de qué variables inciden positiva o negativamente en las tareas de escritura, mayor será la capacidad de controlarlas o de generar tareas estandarizadas que permitan evitar sesgos en los resultados y, en consecuencia, garantizar una evaluación más justa y equitativa. De hecho, la inclusión de estas tareas escritas de respuesta construida está abriendo un amplio campo para la evaluación. Sería importante considerar la posibilidad de emplear tareas integradas, aunque para ello se requiere de recursos tecnológicos que den cuenta de lo que se ha calcado del texto, además de que la escritura dependería de la correcta selección de textos, tarea que incrementa los recursos humanos que se requieren para el diseño de este tipo de tareas. Por añadidura, en caso de que se desee trabajar con esta modalidad, será importante demostrar que éstas en realidad permiten evaluar la habilidad de dar crédito adecuado a las fuentes y de integrar el material citado de forma adecuada. Una de las mayores ventajas de las tareas integradas es que permiten evaluar un abanico más amplio de habilidades (i.e., resumir, parafrasear y sintetizar información).

Debido a las restricciones de tiempo con que se contesta un examen, a diferencia del desarrollo de trabajos académicos en los cuales el tiempo no se encuentra controlado, valdría la pena pensar en otros instrumentos alternativos que les permitan a las universidades evidenciar el proceso de escritura académica de sus prospectos. Por ejemplo, este proceso se podría evidenciar también pidiéndoles a los candidatos no hispanohablantes que traduzcan un trabajo académico o un capítulo de su tesis del ciclo anterior. Otra solución sería pedirles que hicieran resúmenes o comentarios de artículos. El problema de estas dos alternativas es que nunca se podrá garantizar que la versión final fue redactada por los estudiantes. Por esta razón, estas alternativas deberían coexistir con un ECLAA, mas nunca remplazarlo.

4.5.3. LIMITACIONES Y FUTURAS INVESTIGACIONES

Una de las principales limitantes del estudio se debe a la falta de poder sistematizar los diversos géneros textuales que emplean los docentes para evaluar el desempeño académico de sus estudiantes de posgrado. Pese a que la descripción de los docentes respecto de las tareas de escritura que les solicitan a sus estudiantes es detallada, la ambigüedad y falta de precisión respecto del tipo de trabajos solicitados solamente podría esclarecerse mediante un trabajo de análisis textual como el llevado a cabo por Gardner y Nesi (2012). Por ello, estudios posteriores podrían llevarse a cabo mediante la recopilación de textos producidos por estudiantes de posgrado, esto con el fin de identificar con mayor detalle las distintas tipologías textuales usadas dentro del ámbito académico. Estos estudios también podrían ser abordados o complementados a partir de un análisis detallado de las características gramaticales, sintácticas y oracionales a través de la LSF.

Finalmente, una de las principales limitaciones del presente estudio se debe a que no se abordan las variables del candidato ni se cuenta con estudios que evidencien la interacción del sustentante con las tareas de escritura. Por ende, en futuras investigaciones sería importante conocer la opinión de los sustentantes respecto de la tarea propuesta; es decir, mediante entrevistas rememoradas o cuestionarios abiertos o cerrados, indagar sobre posibles problemas o dificultades que pudieran haber encontrado con la temática, las gráficas, la consigna de la tarea, la relevancia del tema a escribir, el tiempo otorgado y otros factores externos a la prueba que pudieran incidir en la resolución de ésta (espacio, mobiliario, orden de la tarea con respecto a las otras pruebas, etc.).

4.6. CONCLUSIONES

El análisis del ámbito para el cual se evalúa resulta una pieza fundamental dentro del proceso de concepción y validación de un examen. En el caso de la escritura académica, conocer las prácticas de evaluación docente y las habilidades de escritura necesarias para que los estudiantes sean académicamente competentes es de vital importancia para justificar la existencia de un examen y su diseño. En el caso de los ECLAA la descripción del ámbito representa una tarea compleja que requiere del apoyo institucional para poder generar bases de datos lo suficientemente grandes y representativas de la población estudiantil. Sólo de esta manera se podrían conocer las prácticas de evaluación docente con detalle y en cada una de las áreas del saber. Como se pudo apreciar en este primer estudio, las formas de evaluar el

conocimiento de los estudiantes difieren entre sí. Algunas dan mayor prioridad a la evaluación mediante exámenes, mientras que en otras se prefiere evaluar mediante trabajos o proyectos. En relación con la importancia de las habilidades de escritura, resalta la importancia de aquellas orientadas a la construcción del conocimiento y a la competencia comunicativa escrita (enmarcada dentro los parámetros del modelo de Canale y Swain, 1980). Por esta razón, valdría la pena explorar la inclusión de tareas integradas (i.e., de lectura para la escritura) en los ECLAA.

CAPÍTULO 5

ESTUDIO 2: LA EVALUACIÓN Y LA GENERALIZACIÓN DE LOS RESULTADOS DE LA PE DEL EXEELAA

5.1. INTRODUCCIÓN

En este estudio se abordaron dos inferencias del argumento de interpretación de la PE del EXEELAA: la evaluación y la generalización de los resultados. La evaluación hace referencia al hecho de que las observaciones del desempeño en las tareas son evaluadas mediante un proceso de calidad que garantiza que los resultados que se le otorgan a los examinados son justos y confiables para tomar decisiones. La generalización se puede demostrar de diferentes formas; por ejemplo, si los estudiantes obtuvieran resultados semejantes en versiones paralelas del mismo examen, o si los resultados pudieran ser comparables con otro tipo de exámenes diseñados con el mismo fin. No obstante, la generalización también se podría determinar en un mismo examen al comparar los resultados que los examinados obtienen en las distintas tareas destinadas a medir una habilidad en particular. Por último, la generalización también se puede evidenciar al comparar los resultados que otorgan los evaluadores a un examinado en una misma prueba de desempeño. Dependiendo del tipo de análisis que se desee llevar a cabo, las inferencias de evaluación y generalización se pueden encontrar estrechamente relacionadas.

En el presente estudio, la inferencia de evaluación y de generalización se determinó a partir de la calidad en el proceso de evaluación: el funcionamiento de la rúbrica y la consistencia de los evaluadores. En primer lugar, se hará una descripción detallada del tipo de rúbricas que mayoritariamente se emplean en los ECLAA; y posteriormente, se describirá el proceso de construcción y de validación de estas escalas de evaluación. Más adelante, se abordará el tema de los evaluadores, se describirán aspectos importantes dentro del proceso de entrenamiento de éstos (i.e., la capacitación y el monitoreo), sus variables personales (i.e., la cognición y sus antecedentes) y los efectos negativos que despliegan al momento de evaluar (i.e., el efecto halo y la centralidad) debido en parte a la interacción que se genera entre el evaluador y el texto al usar la rúbrica. Para poder respaldar estas inferencias, se empleó un estudio cuantitativo inferencial –utilizando el programa estadístico FACTES– mediante el análisis many-facet Rasch measurement (MFRM) para evidenciar la calidad del proceso de evaluación. Este programa permite establecer los índices de calidad y de severidad de las distintas facetas que inciden en

el proceso de evaluación (i.e., el examinado, la tarea, el evaluador y las categorías de la rúbrica). A continuación, se presenta el marco teórico subyacente a este segundo estudio. Posteriormente se describirán los métodos e instrumentos empleados, se presentarán los resultados, seguidos por las discusiones y, finalmente, las conclusiones de este estudio.

5.2. MARCO TEÓRICO

Debido a que este segundo estudio contiene dos inferencias: la evaluación y la generalización, primero se hará una descripción detallada de las rúbricas y posteriormente se abordará a los evaluadores.

5.2.1. LAS RÚBRICAS EMPLEADAS PARA EVALUAR LAS PE

La evaluación de las PE mediante rúbricas comenzó a crecer rápidamente en EE.UU. a partir de los 70 con la necesidad de contar con una evaluación más justa y equitativa para los sustentantes. De hecho, hasta finales de los 90, los estudios estadísticos que se habían conducido en relación con éstas eran relativamente escasas (Jonsson y Svingby, 2007). A las rúbricas también se les conoce como baremos, pautas de corrección y como escalas de evaluación (Hamp-Lyons, 2002). La rúbrica es un instrumento que se emplea para evaluar pruebas de respuesta construida y debe responder a las bases teóricas a partir de las cuales se diseñó un examen. Sin embargo, a partir de su uso, se desprenden dos problemáticas importantes a considerar: la rúbrica deberá medir lo que pretendemos medir, y los evaluadores, mediante capacitación, deberán ser capaces de usarla de manera estandarizada.

De hecho, East (2009) menciona que “gran parte de la variabilidad de los resultados se debe a la idiosincrasia de las sesiones de entrenamiento, a los evaluadores y a las complejas y peculiares impresiones que influyen las decisiones de la evaluación, independientemente de cómo está redactada la rúbrica (p. 90). En efecto, una de las críticas más asiduas a las rúbricas es que éstas nunca lograrán ser usadas de manera uniforme por los evaluadores. Incluso, uno de los mayores problemas que se evidencia en su uso es que un mismo evaluador puede evaluar un texto de forma diferente en dos ocasiones dentro de una misma sesión de evaluación.

En el caso del diseño de la rúbrica, Becker (2011), Knoch (2007) y Weigle (2002) hacen una extensa revisión de los distintos tipos que existen, así como de las ventajas y desventajas de la implementación de uno u otro tipo. Sin embargo, más que el tipo de rúbrica que se emplea,

una cuestión fundamental consiste en de qué forma se usa por el evaluador y con qué propósito. De hecho, uno de los intereses que ha girado en torno a la validación de las rúbricas se debe a que la implementación de éstas en la evaluación no garantiza per se una evaluación más eficaz (Rezaei y Lovorn, 2010). En el estudio que estos últimos investigadores condujeron, los evaluadores no mostraron mayor consistencia en sus resultados al contrastar las evaluaciones con y sin rúbrica. Los autores mencionan que cuando los evaluadores no son bien entrenados en cómo usar las rúbricas y a qué niveles de desempeño corresponden cada uno de los niveles contemplados en los criterios, su uso no hace ninguna diferencia.

No obstante, una de las grandes ventajas de las rúbricas (analíticas u holísticas) dentro de la evaluación a gran escala, alto impacto o incluso dentro del salón de clases, es que el estudiante o sustentante de un examen puede ser informado claramente sobre cuáles son los aspectos que se considerarán para evaluar su desempeño. En este caso, estamos hablando de PE, pero también podrían ser de pruebas orales. Por lo tanto, es necesario informar a las distintas partes interesadas (estudiantes, diseñadores de cursos o quienes toman decisiones) sobre cómo se evalúa y qué aspectos se consideran importantes.

5.2.1.1. EL TIPO DE RÚBRICAS

De manera muy general, las rúbricas se pueden dividir en dos categorías: holísticas y analíticas. Las rúbricas holísticas son aquellas en las cuales se utiliza un sólo puntaje para evaluar el desempeño de un candidato; en este caso, en una tarea escrita. Las rúbricas analíticas, por su parte, constan de una serie de rasgos o criterios que el evaluador considera importantes dentro del proceso de evaluación. Mediante el uso de estas escalas de evaluación se puede obtener también un puntaje global o fragmentado cuando cada uno de los rasgos aporta información que puede ser empleada, por ejemplo, para el diseño de cursos. Weigle (2002) asevera que los argumentos a favor o en contra de uno u otro tipo de rúbrica son vastos; así que, más que establecer un juicio sobre qué tipo de rúbrica es mejor, se hará una descripción de éstas con sus bondades y perjuicios dependiendo del propósito para el cual se las emplea. Sin embargo, Lai, Wolfe y Vickers (2015) mencionan que los procesos cognitivos al emplear una u otra rúbrica son diferentes, lo cual repercute en el evaluador y también en la recopilación de datos.

5.2.1.1.1. LAS RÚBRICAS HOLÍSTICAS

Este tipo de pautas de corrección se emplean frecuentemente en exámenes administrados a gran escala (Becker, 2011; Knoch, 2007b; Weigle, 2002). La razón fundamental de su uso en dichos exámenes radica en que es barata, precisa y se asume que es práctica al momento de ser usada (Jonsson y Svingby, 2007), pues el corrector hace una primera lectura del texto que desea evaluar y a partir de una primera impresión considera de forma global los distintos aspectos que se han considerado importantes dentro de la evaluación final. Rezaei y Lovorn (2010) agregan que éstas se orientan más al producto terminado. Las rúbricas holísticas generalmente incluyen rangos que varían entre seis y nueve distintos niveles (TOEFL iBT, IELTS, Test DaF, GRE). De hecho, Alderson et al. (1995) recomiendan que las rúbricas no cuenten con más de 7 niveles o bandas, pues resulta difícil para los evaluadores hacer una fina distinción entre éstos. Una de las críticas cuando se incluyen demasiados niveles es que los evaluadores no logran discernir claramente entre distintos niveles de actuación. Sin embargo, Attali, Lewis y Steier (2012) encontraron que cuando la rúbrica es más detallada, los evaluadores más experimentados podían emplearla mejor, aun cuando no hubieran tenido capacitación adicional en el uso de una nueva rúbrica. Asimismo, agregan que la confiabilidad de los resultados incrementa al tener un espectro de evaluación más amplio.

Becker (2011), Lee, Gentile y Kantor (2008) y Weigle (2002) comentan que otra de las ventajas de las rúbricas holísticas es que el corrector se enfoca en las fortalezas de la escritura del sustentante y no en las deficiencias o errores de éste. Por ende, cuando un examinador revisa un texto, compensa favorablemente el desempeño de un candidato cuando su actuación es adecuada, aunque presente algunos errores ortográficos o de puntuación, por ejemplo. Según Weigle (2002), el corrector pone mayor énfasis en el contenido y la manera mediante la cual el sustentante lo transmite, y no en la forma ni en la mecánica. Una última característica positiva de las rúbricas holísticas, según Barkaoui (2007), radica en que el evaluador puede incluir en su evaluación criterios que no se encuentran dentro de las especificaciones establecidas, lo cual le confiere cierta libertad al evaluador para emplear la escala de manera flexible en casos en los cuales ciertos criterios no han sido incluidos en ésta. Sin embargo, si bien las escalas de evaluación holísticas permiten esta flexibilidad, una limitante es que se reduce la consistencia de los evaluadores tanto intra como interevaluador (Barkaoui, 2010b). Naturalmente, si esto sucediera, entonces también el significado de los resultados se alteraría.

Contrariamente, dentro de las desventajas de las rúbricas holísticas, East (2009), Rezaei y Lovorn (2010) y Weigle (2002) afirman que los evaluadores se ven altamente influenciados por aspectos irrelevantes de la escritura como la caligrafía o la longitud del texto. Investigadores como Lee et al. (2008) encontraron que la longitud de los textos correlacionaba altamente con el puntaje otorgado por las evaluaciones holísticas y con ciertos criterios de las escalas de evaluación analíticas, como el desarrollo de ideas (cfr. Attali et al., 2012). Aunado a lo anterior, cuando se hace una evaluación holística, no es fácil identificar el criterio que ha seguido el evaluador para identificar el nivel o la banda del candidato, ya que cada una de estas bandas contiene diversos elementos considerados como importantes, pero no necesariamente el examinando los reúne todos (Lee et al., 2008). Es decir, por ejemplo, un evaluador puede compensar ciertas deficiencias en la gramática y en la organización de un texto que contiene ideas muy acertadas y propositivas, y otorgar el mismo puntaje que a otro texto en donde las ideas no son muy claras, pero la organización es coherente y la gramática muy precisa y variada. De hecho, East (2009) y Knoch (2007b) comentan que debido a que los evaluadores se ven en la necesidad de otorgar solamente una nota final, es común que mayoritariamente dependan de una o dos de las características enunciadas por el descriptor.

Otro punto en desventaja de las rúbricas holísticas se debe a que éstas no proporcionan información diagnóstica sobre el desempeño del sustentante (Knoch, 2007, 2009, 2011; Lee et al., 2008). Esto es importante, pues uno de los fines últimos de la evaluación es que conlleve consecuencias positivas para las partes interesadas. En este caso, tratándose de los sustentantes y docentes, éstos no pueden identificar qué es importante mejorar en cuanto a su escritura, pues la rúbrica sólo da un puntaje global en el cual se enmarcan diversas habilidades que no necesariamente son iguales. Asimismo, en el caso de que el objetivo de la evaluación sea clasificar a los estudiantes en distintos grupos de escritura, la rúbrica holística no puede servir como diagnóstico para la elaboración de planes de estudio o mejoramiento de la escritura académica de los estudiantes menos proficientes.

Finalmente, pese a que tanto las rúbricas holísticas como las analíticas requieren de capacitación por quienes las emplearán para evaluar –debido a que las holísticas son más difíciles al momento de usar, ya que el desempeño en diversos criterios puede caer en varios niveles o bandas– se requiere de mucha mayor práctica, capacitación y profesionalización por parte de los correctores.

5.2.1.1.2. LAS RÚBRICAS ANALÍTICAS

Las pautas de corrección analíticas constan de una serie de rasgos que se evalúan de forma independiente. Rezaei y Lovorn (2010) mencionan que éstas se encuentran más bien centradas en el proceso que en el producto final. Una de las principales ventajas de su uso es que proveen información detallada respecto del desempeño del candidato; información que es útil, por ejemplo, cuando el fin de la evaluación es diagnóstico (Knoch, 2007, 2009). De igual manera, resultan útiles cuando uno de los fines radica en la implementación de cursos que fomenten el desarrollo de las habilidades de escritura (Hamp-Lyons, 1995; Jonsson y Svingby, 2007; Knoch, 2007; Weigle, 2002; Zhao, 2012).

Otra de las ventajas del uso de las rúbricas analíticas consiste en el incremento en la confiabilidad (Hamp-Lyons, 1991b; Jonsson y Svingby, 2007; Knoch, 2009). De hecho, East (2009) añade que el proceso de evaluación de los correctores es más consistente con este tipo de rúbricas. Como se expuso anteriormente, en el caso de las rúbricas holísticas se emplean pocas bandas o niveles de desempeño; no obstante, debido a que la rúbrica analítica es detallada, cada rasgo podría evaluarse de igual manera en cinco o siete niveles de actuación, multiplicándose así considerablemente el puntaje total y, por lo tanto, la probabilidad de adyacencia entre un evaluador y otro se incrementa. Como Knoch (2007) y Weigle (2002) mencionan, las rúbricas analíticas son más confiables, pues al igual que una prueba de opción múltiple incrementa su confiabilidad cuando se agregan más reactivos, igual sucede con las rúbricas analíticas. Knoch (2007) menciona que otra de las ventajas de las rúbricas analíticas es que éstas pueden contar con diferentes niveles o bandas para cada uno de los criterios, ya que no siempre es pertinente que todas las categorías cuenten con el mismo número de bandas. Contrariamente, uno de los riesgos consiste en que el puntaje final sea similar entre dos evaluadores, pero los puntajes internos no.

En el caso de la evaluación de segundas lenguas, dado que los sustentantes pueden presentar diversos niveles de dominio de escritura en la lengua meta, contar con una rúbrica analítica favorece una evaluación más equitativa (Weigle, 2002), ya que el candidato puede obtener puntajes altos en algunos rubros y compensarlos con puntajes bajos en otros. Por último, otra de las ventajas cuando se diseñan rúbricas analíticas es que se puede otorgar mayor peso a un rubro en comparación con otros (Knoch, 2011; Weigle, 2002). Así pues, si diseñamos una tarea en la cual el foco consiste en el desarrollo de un argumento dado a partir de diversas posturas, se le puede dar mayor peso al desarrollo de las ideas que a la ortografía, por ejemplo.

Saxton, Belanger y Becker (2012) arguyen que cuando se trata de PE que ponen en juego el pensamiento crítico y las subhabilidades que de éste se derivan, debería optarse por una rúbrica analítica. Esto debido a que, según los autores, cuando se emplea una rúbrica holística, se asimilan y homogenizan los distintos criterios para dar un solo resultado. Al hacer este proceso de asimilación de criterios, se pierde información detallada de las diferentes habilidades de escritura que posee un examinado. Aunado a esto, cuando se emplea una rúbrica analítica, se puede establecer claramente qué es lo que se está midiendo.

No obstante, una de las principales desventajas de su adopción subyace en el tiempo que se requiere para evaluar detalladamente cada uno de los aspectos identificados como importantes de acuerdo al propósito del examen. En este sentido, cuando se trata de exámenes administrados a gran escala, resulta impráctico su uso debido al gran volumen de textos que se necesitan evaluar. Adicionalmente, el tiempo empleado en su uso va de la mano con los costos requeridos (Lee et al., 2008), razón por la cual resulta impráctica en dicho contexto. Sin embargo, en el caso de los exámenes de alto impacto administrados a pocos sujetos, su uso podría resultar adecuado.

Otra de las críticas que se les han hecho a las rúbricas analíticas es que frecuentemente se emplean criterios vagos para clasificar los distintos niveles de actuación. Por ejemplo, si se desea evaluar el léxico, utilizar adjetivos como *adecuado*, *suficiente*, *mínimo* no dan cuenta en realidad de a qué hacen referencia cada uno de éstos, volviéndose así subjetiva la evaluación y de carácter impresionista (Knoch, 2011; Lee et al., 2008). Aunado a esto, cuando los rasgos no son claramente descritos, los evaluadores presentan problemas para aplicar el criterio y recurren a la subjetividad, al efecto halo¹ o a la centralidad para la evaluación de dicho rasgo (Eckes, 2009a; Knoch, 2007; Lai et al., 2015; Myford y Wolf, 2003, 2004). Por ello, autores como Knoch (2009) y Rezaei y Lovorn (2010) mencionan que es imprescindible contar con rúbricas analíticas lo más detalladas posible para evitar que los evaluadores empleen recursos impresionistas para evaluar los textos.

¹ El efecto halo hace referencia a una evaluación subjetiva que se ve influenciada por los criterios anteriores de la rúbrica. Lai et al. (2015) se refieren a éste como el “halo ilusorio”, pues mencionan que éste puede ser introducido en varios contextos: a) cuando una impresión general del texto distorsiona la manera en la que se evalúa cada categoría, b) cuando la percepción de categorías sobresalientes distorsionan la evaluación de las otras, y c) cuando el evaluador no es capaz de discriminar entre los puntajes de las diferentes categorías, ya sea porque hay ambigüedades en la rúbrica o no tuvo suficiente entrenamiento (p.105).

El Cuadro 5.1 resume algunos aspectos importantes a considerar al momento de optar por una rúbrica analítica u holística. Como se mencionó al inicio de este apartado, no se pretende dar cuenta de qué rúbrica es mejor, sino simplemente describir cuáles son los beneficios y los escollos de emplear una u otra. Becker (2011, p. 128) apunta que, “mientras que hay copiosas investigaciones que comparan el potencial de las ventajas y desventajas al emplear un cierto tipo de rúbricas para evaluar, hay muy poca evidencia en las investigaciones sobre las decisiones tras la selección de éstas”. Será importante considerar, entonces, el propósito para el cual deseamos evaluar, la cantidad de sustentantes que presentarán dicha prueba y los recursos económicos y humanos con los que cuenta la institución para evaluar los textos.

Cuadro 5.1: Ventajas y desventajas de las rúbricas holísticas y analíticas
(Elaborado por el investigador)

RUBRICAS HOLÍSTICAS	
Ventajas	Desventajas
<ul style="list-style-type: none"> - La evaluación de los textos es más rápida. - Orientada hacia el producto terminado o al logro de la tarea. - El foco de la evaluación radica en las fortalezas y el contenido, no en las debilidades o en la mecánica. - Flexible para evaluar aspectos no incluidos en los criterios. 	<ul style="list-style-type: none"> - Menor confiabilidad en los resultados. - No se puede saber cómo llegó el evaluador al resultado que otorgó. - Selección de un par de criterios para evaluar. - Evaluación impresionista de rasgos superfluos: longitud del texto, caligrafía. - Requieren de mayor capacitación y experiencia de uso por parte de los evaluadores. - No proporcionan información diagnóstica.
RUBRICAS ANALÍTICAS	
Ventajas	Desventajas
<ul style="list-style-type: none"> - Evaluación de criterios de forma independiente. - Centradas en el proceso más que en el producto. - Proveen información detallada de las habilidades de escritura del sustentante. - Mayor confiabilidad. - Mayor consistencia en las evaluaciones globales, aunque no necesariamente en las internas. - Benéficas para evaluar segundas lenguas, pues los perfiles de los sustentantes son desiguales. - Posibilidad de otorgar mayor peso a ciertos rubros de la rúbrica. - Son más fáciles de emplear por evaluadores novatos. 	<ul style="list-style-type: none"> - Su diseño requiere mayor tiempo. - Se requiere mayor tiempo para evaluar. - Resultan poco prácticas para evaluar a gran escala. - La ambigüedad o falta de claridad en los descriptores puede favorecer el efecto halo, la centralidad o la evaluación de tipo impresionista.

Dentro de los estudios diseñados para comparar estadísticamente el funcionamiento de rúbricas analíticas y holísticas, destacan el de Barkaoui (2007) y Lee et al. (2008). En el caso del

primero, contrariamente a lo que el investigador esperaba, los resultados arrojados al emplear una rúbrica holística resultaron más confiables y consistentes que los de su contraparte analítica. No obstante, apunta el autor, esto no significa que los evaluadores se encuentren en consenso, sino que puede ser que múltiples factores inciden, de manera distinta, para que lleguen al mismo resultado. Por esta razón, los estudios sobre el proceso cognitivo del corrector merecen especial atención.

Finalmente, destaca el amplio estudio de Lee et al. (2008) para ver de qué forma correlacionaba la rúbrica holística empleada para evaluar las PE del TOEFL iBT y una rúbrica analítica que se construyó con fines comparativos. Los autores destacan que los seis criterios seleccionados en la rúbrica analítica correlacionaban significativamente entre ellos, pero además con la rúbrica holística. De éstos, los que mejor correlacionaron con la rúbrica holística fueron: el desarrollo, el vocabulario y la variedad de oraciones. También la correlación entre la organización y el uso de la lengua correlacionaron adecuadamente. El único criterio que denotó una baja correlación fue el de aspectos mecánicos (errores gramaticales, ortográficos y de puntuación). Los autores concluyen que los resultados holísticos reflejan asertivamente la calidad del contenido y del lenguaje utilizado en los ensayos argumentativos del TOEFL iBT, mientras que los analíticos no fueron capaces de discernir con mayor grado de detalle la habilidad de los sustentantes. Por ende, los autores argumentan que el uso de una rúbrica analítica no se justifica bajo el supuesto argumento que su uso incrementa la confiabilidad en las evaluaciones a gran escala.

En el caso de la PE del EXELEAA, se optó por una rúbrica analítica por varios motivos. En primer lugar, dado que el número de sustentantes que presentan el examen anualmente es bajo (menos de 100), resulta importante que los correctores tomen en cuenta todos los criterios que se consideran esenciales para evaluar los textos. En este sentido, si se empleara una rúbrica holística, el espaciamiento entre exámenes y el bajo número de textos a evaluar por corrector acrecentaría la falta de consistencia en las evaluaciones. En segundo lugar, dado que el número de sustentantes es bajo, el costo que genera el uso de una rúbrica analítica no representa un problema. Finalmente, además de que se puede obtener un puntaje más fino por parte de los correctores, se pueden llevar a cabo estudios estadísticos que permitan evidenciar problemas puntuales de cada uno de los evaluadores o de las categorías de la rúbrica, con lo cual se puedan implementar mejoras en la escala de evaluación y también sesiones de entrenamiento y estandarización entre los correctores.

5.2.1.2. EL PROCESO DE CONSTRUCCIÓN DE UNA RÚBRICA

El origen de las rúbricas puede ser obscuro y poco claro, pues en muchas ocasiones se adaptan de otras que han sido creadas originalmente para fines distintos a los cuales se pretende evaluar. De hecho, Becker (2011) identificó que la gran mayoría de las instituciones educativas de los EE.UU. que fue encuestada mencionó haber adaptado rúbricas del TOEFL iBT para evaluar el desempeño de escritura de sus estudiantes. En términos generales, las rúbricas deberían responder al constructo a partir del cual se diseñó el examen; es decir, éstas deberían encontrarse alineadas a la descripción de las habilidades que se pretende evaluar de acuerdo al propósito del examen (Knoch, 2011; Montee and Malone, 2014). Sin embargo, muchas veces la construcción de la rúbrica se hace de forma intuitiva. Knoch (2007, 2009), siguiendo la línea de Fulcher (2003) describe tres aproximaciones para construir una rúbrica: basadas en la intuición, en una teoría o de forma empírica.

En el primer caso, la autora comenta que las rúbricas intuitivas son aquellas que se construyen a partir del juicio de expertos y son de tipo experiencial. Para diseñar estas rúbricas, generalmente se le pide a un grupo de evaluadores experimentados que corrijan una cantidad considerable de textos representativos del tipo de examen que se está diseñando o revisando. El fin de esta evaluación es que los jueces determinen cuáles son las características principales que emergen de los textos y que podrían ser incorporadas en la rúbrica. Generalmente, éstas se construyen tomando como base otras rúbricas preexistentes para el examen en cuestión o de otras rúbricas de exámenes similares. No obstante, uno de los riesgos de construir rúbricas de este tipo se debe a que los criterios frecuentemente empleados se describen con vaguedad. Aunado a lo anterior, Becker (2011) asevera que este proceso es demasiado tardado y también requiere de una inversión considerable de recursos económicos y humanos.

En el caso de las rúbricas que se sustentan en un modelo de escritura, éstas representan una simplificación de un fenómeno que es complejo, como la escritura, que además no se encuentra circunscrita a ninguna teoría que permita dar cuenta de cómo es el proceso de escritura (Zhao, 2012). Existen modelos de competencia comunicativa que se han elaborado para tratar de describir este proceso (Canale y Swain, 1980; Bachman y Palmer, 1996, 2010; McNamara, 1996; Weir, 2005), pero su utilización y adaptación para evaluar con un fin específico no es sencillo, pues no siempre es posible evaluar todos los componentes que se encuentran incluidos en un modelo de escritura. Por ejemplo, en el estudio que Zhao condujo para construir una rúbrica analítica basándose en el modelo interaccional de voz de Hyland (2008), la autora

incluyó los nueve componentes que Hyland incluye en su modelo, más dos que la autora consideró importantes incluir; de ahí que, la rúbrica analítica constara de 11 categorías que podían ser evaluadas en una escala del 0-4. La autora utilizó 200 textos de la tarea independiente del TOEFL iBT (redacción de ensayo argumentativo) y seis evaluadores experimentados y con previa capacitación en el uso de la rúbrica se dieron a la tarea de evaluar doscientos ensayos del TOEFL iBT. Sin embargo, la autora encontró que cuatro categorías no podían ser evaluadas por la rúbrica, pues los textos no evidenciaban el uso de todos los componentes del modelo de Hyland. En suma, el modelo no podía ser empleado en su totalidad para evaluar los textos académicos del TOEFL iBT.

Del estudio de Zhao (2012) se puede concluir que aun cuando una rúbrica analítica sea construida con base en un modelo teórico, algunas de las categorías pueden carecer de relevancia, dependiendo de en qué contexto se la emplee. En este caso, tratándose de textos cortos del TOEFL iBT, cuya longitud es de aproximadamente 300 palabras, la autoridad en la voz de quien escribe dista mucho de otros registros escritos extensos. Por ende, resulta indispensable considerar las características de la tarea escrita en el diseño de una rúbrica.

Por esta razón, Knoch (2007) propone que las rúbricas sean construidas de forma empírica; esto es, a través de procedimientos estadísticos que permitan dar cuenta de la relación que existe entre el comportamiento lingüístico, la tarea y la rúbrica. Sin embargo, el resultado de la extensa investigación de Knoch (2007) para diseñar una rúbrica analítica de tipo empírico denota que los criterios descritos en ésta son difíciles de aplicar por parte de los correctores humanos. La autora, en su estudio sobre validación de rúbricas analíticas, propone que éstas contengan descriptores sumamente detallados del desempeño del sustentante. Sin embargo, este tipo de análisis resulta apropiado para un corrector electrónico, mas no para un humano. Por ejemplo, tras un exhaustivo análisis estadístico, la autora propone que para la precisión, por ejemplo, se le otorgue un 9 a los textos que presentan oraciones sin errores, 8 a los que contienen la mayoría de las oraciones sin errores, 7 cuando aproximadamente $\frac{3}{4}$ partes de las oraciones no contengan errores y así sucesivamente. En cuanto al reparo de los escritos, propone 9 para aquellos que no fueron reparados, 8 a los que contienen no más de 5 autocorrecciones, 7 para los que presentan entre 6-10 autocorrecciones, 6 para el rango de 11 a 15 autocorrecciones, y así sucesivamente. En cuanto al léxico, 8 para textos que contienen más de 20 palabras académicas (éstas se pueden consultar en un apéndice que se le proporciona al evaluador), 7 a aquellos que contienen entre

12 y 20 palabras académicas, 6 para los textos que incluyen entre 5 y 12 palabras académicas, etcétera.

La problemática que se suscita al contar con criterios tan detallados en las rúbricas radica en que la evaluación toma demasiado tiempo a los correctores si ésta es rigurosa de acuerdo con los criterios establecidos. De hecho, el uso de rúbricas tan detalladas implicaría que los correctores funcionaran como máquinas, lo cual conlleva un complejo proceso de evaluación poniendo especial atención a la forma y no tanto al contenido ni a funciones superiores de escritura. El problema fundamental de una rúbrica empírica tan estilizada radica en que los evaluadores pierden la noción del texto como un todo y tienden a fragmentar cada uno de los aspectos a evaluar; sobre todo, poniendo especial énfasis en los errores. En realidad, si los textos son evaluados por humanos, es contradictorio considerarlos como máquinas y, por lo tanto, tampoco podemos esperar correlaciones perfectas en la forma de evaluar, pues como se mencionó al inicio de este apartado, más allá de la rúbrica, las diferencias en la forma de evaluar se encuentran en las características del evaluador y en la manera en la cual han sido capacitados para evaluar los textos. Esto no quiere decir que no existan criterios estandarizados para evaluar las composiciones escritas, pero la visión de un evaluador siempre diferirá hasta cierto grado de lo que otro evaluador observa. En términos generales, si bien es cierto que las rúbricas analíticas son más detalladas, su empleo presenta rasgos holísticos en donde la valoración final resulta un tanto subjetiva, pese a los criterios y lineamientos que previamente hayan sido establecidos.

Lo que sí resulta importante es que la rúbrica se construya a partir de la definición del constructo y del propósito para el cual se desea evaluar, así como de las inferencias y generalizaciones que se desean tomar con base en los resultados. Asimismo, resulta fundamental informar sobre aquellas habilidades de escritura que se evalúan mediante la rúbrica, y en el caso de las analíticas, dar cuenta del funcionamiento de cada uno de los criterios incorporados en ésta. Por ello, dar cuenta del proceso de diseño de la rúbrica resulta tan importante como del proceso de validación de la misma.

5.2.1.3. EL PROCESO DE VALIDACIÓN DE UNA RÚBRICA

En la evaluación de las habilidades lingüísticas a gran escala y/o de alto impacto, cada vez es más y más frecuente la inclusión de pruebas de respuesta construida para evaluar la producción oral y escrita; sin embargo, lo anterior conlleva que se deban conducir estudios que garanticen la consistencia en las evaluaciones (Wind y Engelhard, 2013). Esto se debe a que

variables como la dificultad de la tarea y la severidad o indulgencia de los evaluadores al momento de emplear la rúbrica afectan la confiabilidad.

Rater-mediated assessments offer many advantages over assessments based on multiple-choice items, but these advantages may be illusory if the quality of the ratings yields low levels of reliability and validity for the proposed uses of the rater-mediated assessments (Wind y Engelhard, 2013, p. 298).

Por lo anterior, sin importar si una rúbrica ha sido construida a priori o de forma empírica, es importante validar el instrumento con el objeto de determinar su consistencia y demostrar si, efectivamente, éste mide lo que en realidad pretendemos medir. Antes de revisar los diversos procesos de validación que pueden seguirse, es menester describir las rúbricas que han sido construidas a priori o de forma empírica. En caso de contar con una teoría lingüística subyacente al constructo, es posible establecer de manera preliminar cuáles serán aquellos rasgos que serán evaluados mediante la rúbrica. Una de las características del proceso de diseño de rúbricas es su carácter iterativo (Weigle, 2002), pues es innegable que la revisión de los instrumentos debe ser una tarea constante por parte de quienes la emplean.

Las investigaciones diseñadas para validar rúbricas son diversas. En primer lugar, se describirán dos de ellos que se caracterizan porque el contenido de la rúbrica se valida por un grupo de expertos, en este caso de escritura académica, (Jonsson y Svingby, 2007). Por una parte, tenemos aquellos estudios de carácter empírico que han tratado de identificar qué secciones de una rúbrica analítica diseñada para el ámbito académico son importantes de acuerdo a las habilidades de escritura requeridas para que un estudiante universitario sea académicamente competente (Becker, 2011; Rosenfeld et al., 2004).

En el primer caso, Becker (2011) condujo un estudio con 43 directores de universidades en EEUU para que éstos identificaran qué secciones típicas de las rúbricas (precisión, contenido, coherencia, complejidad, gramática, uso de la lengua, organización, sintaxis/estructura y vocabulario) eran importantes para que se evaluara la escritura en los programas universitarios. En primer lugar, el autor encontró, aunque no de forma generalizada, que las rúbricas holísticas eran preferidas a las analíticas, particularmente cuando se trata de evaluar las competencias de los estudiantes. Entre algunas de las razones dadas fueron: practicidad y rapidez. Sin embargo, quienes consideraron que las analíticas eran mejores, argumentaron que eran mucho más informativas. Estos hallazgos van de la mano con las ventajas y desventajas que se han manifestado en la literatura respecto del uso de uno u otro tipo de rúbrica. Aunado a lo anterior,

las categorías de las rúbricas que los directores destacaron como más importantes fueron: precisión, organización y vocabulario.

En otro estudio similar, Rosenfeld et al. (2004) identificaron previamente qué habilidades eran consideradas importantes para el ámbito académico (ver Capítulo 4, Estudio 1). Posteriormente, un grupo conformado por cinco especialistas cotejó las habilidades que los docentes consideraban importantes para el éxito académico y cuáles de éstas eran evaluadas por la rúbrica. Los resultados del proceso de vinculación mostraron que 25 de los 39 enunciados de tareas de escritura contenidos en la encuesta eran evaluados por la rúbrica analítica diseñada para el examen *Graduate Record Examination* (GRE)

Además de la validación de las rúbricas a partir del juicio de expertos, existen también métodos estadísticos que permiten evidenciar el funcionamiento interno de una rúbrica analítica. Para ello, se puede emplear el análisis factorial, mediante el análisis del componente principal, *Principal Component Analysis* (PCA) o el factor de ejes principales, *Principal Axis Factoring* (PAF). Estos procedimientos se pueden emplear cuando existe una teoría subyacente a la construcción de la rúbrica o no. Cuando las rúbricas analíticas se construyen a priori con sustento en una teoría lingüística, no se sabe con exactitud si las categorías para evaluar la producción escrita se verán reflejadas en la tarea solicitada al sustentante; de ahí que, el análisis factorial de reducción de datos sirva como instrumento para identificar un pequeño número de factores a partir de múltiples variables observadas que se correlacionan entre sí. Este tipo de análisis resulta útil cuando se miden constructos que no pueden ser observables en la vida real, como es el caso de las habilidades de escritura.

Autores como Knoch (2007, 2009) y Zhao (2012) han empleado este procedimiento para determinar los factores principales que mejor explican el fenómeno de la escritura. Retomando el estudio mencionado en el apartado anterior que Zhao (2012) condujo para construir y validar una rúbrica basada en el modelo interaccional de voz (Hyland, 2008), la autora empleó el PCA con el fin de explorar los componentes que definían el constructo de la voz en los ensayos del TOEFL iBT. Tras el análisis, la autora identificó tres componentes principales de la rúbrica a los cuales denominó: la presencia y claridad de ideas en el contenido, la manera de presentar las ideas, y la presencia del escritor y lector. Cada una de las siete categorías (pues cuatro de ellas fueron excluidas) se agrupó en uno de los tres componentes descritos.

Knoch (2009) condujo un estudio mixto (estudios cuantitativos y cualitativos) para contrastar dos tipos de rúbricas analíticas: una construida a priori y otra desarrollada de forma empírica. En una primera fase, se revisaron 600 ensayos. Mediante un minucioso análisis estadístico del comportamiento de los sustentantes en diversos criterios que de antemano se consideraron importantes para evaluar la escritura académica, se creó una nueva rúbrica. Posteriormente en una segunda fase, tras previo entrenamiento mediante un manual, 10 evaluadores revisaron 100 ensayos más empleando la rúbrica antigua (construida a priori), y dos meses después, la segunda rúbrica (construida de forma empírica). Una vez revisados los ensayos, se corrieron dos estudios estadísticos, uno empleando el análisis factorial mediante el PCA y otro más a través del análisis de MFRM. Más adelante, se entrevistó a los evaluadores con el fin de saber sus percepciones en el uso de ambas rúbricas. El objetivo de comparar el funcionamiento de ambas rúbricas se debía a que la primera versión contaba con descriptores que empleaban adjetivos vagos e imprecisos como: “adecuado”, “suficiente”, “excepcional”, etc., además que ciertos criterios estaban en conflicto por estar representados en dos o más rubros de la escala.

Mediante al análisis factorial del PCA, la autora identificó que a la nueva rúbrica subyacían seis factores que explicaban mayoritariamente la varianza, mientras que en la primera rúbrica sólo uno. En relación con las entrevistas, la autora identificó que cuando los evaluadores empleaban descriptores vagos, como los empleados en la primera rúbrica, estos evaluadores empleaban otras estrategias para compensar las deficiencias de la rúbrica: asignar una nota global, evaluar mediante el efecto halo y descartar el descriptor. En términos generales, los autores que se inclinaron en favor de la segunda rúbrica, mencionaron que era más útil porque era más detallada y clara, aunque no todos percibieron el cambio de rúbrica favorable, puesto que evidenciaron que la nueva rúbrica le restaba flexibilidad al procedimiento de evaluación y dejaba fuera el sentido crítico del evaluador. Sin embargo, Knoch (2009) destaca que los evaluadores deben estar muy conscientes del propósito de la evaluación en las sesiones de entrenamiento.

Finalmente, Knoch (2009) y Rezaei y Lovorn (2010) comentan que el tema de las rúbricas debe tomarse con cautela. Aun cuando es cierto que una rúbrica bien diseñada y con ejemplos prototípicos debería ayudar a la evaluación, ¿qué pasa con aquellas que son adaptadas de internet e implementadas sin ningún entrenamiento? El uso incorrecto de una escala de evaluación podría incluso empeorar la evaluación cuando ésta no se emplea con la suficiente

capacitación y entrenamiento. Más aún, East (2009) agrega que las rúbricas no pueden ser intercambiables entre distintos idiomas. El autor pone como ejemplo que los parámetros para evaluar una oración simple en inglés (con complementos directos e indirectos) no puede ser la misma que para el alemán, dado que en esta última lengua, los complementos se declinan según el caso acusativo y dativo, además de la complejidad de identificar entre tres géneros: masculino, femenino y neutro, los cuales, a su vez, pueden formar hasta siete plurales distintos. Mientras tanto, en el inglés no existen casos en complementos, ni géneros y solamente una forma del plural. Por estas razones, el diseño e implementación de las rúbricas requiere de especial cuidado y no pueden simplemente trasladarse de un contexto a otro o traducirse de una lengua a otra.

5.2.2. LOS EVALUADORES DE PE

La consistencia en la evaluación de pruebas de desempeño se ha convertido en el foco de atención de quienes diseñan PE en exámenes de lenguas, puesto que la correcta aplicación del instrumento de medición (en este caso la rúbrica) resulta indispensable para garantizar la confiabilidad de un examen y, por lo tanto, salvaguardar la validez de este tipo de pruebas. Enright y Quinlan (2010) mencionan que el gran problema con los evaluadores humanos se debe a la manera en la cual interpretan y aplican la rúbrica, lo cual pone en detrimento la calidad de la evaluación y propicia la obtención de resultados espurios. Por esta razón, es indispensable que este tipo de pruebas cuenten con dobles jueces y el mayor número de interacciones posibles², ya que sólo de ésta manera se podrá establecer el índice de confiabilidad entre las evaluaciones y determinar hasta qué punto los resultados de los evaluadores son consistentes. De hecho, Wang y Yao (2013) mencionan que existe fuerte evidencia que los evaluadores humanos difieren en su índice de severidad, lo cual conduce a sesgos y errores sistemáticos causados por la severidad o indulgencia del evaluador. Los autores mencionan que resulta importante asegurarse que la asignación de los escritos a los evaluadores se haga de manera aleatoria. Los autores encontraron que los sesgos entre evaluadores incrementaban cuando las asignaciones de las evaluaciones no se hacían de forma aleatoria.

² Por interacción se hace referencia al diseño de una matriz que permita comparar, en la mayor medida posible, a todos los evaluadores. Si el evaluador 1 siempre evalúa junto con el evaluador 2 ó 3, pero no con el 5, 6... *N*, entonces no se puede lograr una red que permita evidenciar óptimamente el índice de severidad de cada uno de los evaluadores. De hecho, como menciona Knoch (2007), se debe buscar que los textos sean evaluados por al menos dos correctores.

En el caso de los exámenes de lengua con fines académicos, existen diversas maneras de llevar a cabo el proceso de corrección de PE. Por ejemplo, en el caso del IELTS, solamente interviene un evaluador, aunque existen monitoreos periódicos y aleatorios por parte de evaluadores experimentados que permiten identificar qué evaluadores están siendo demasiado severos o indulgentes al momento de evaluar. En otros casos, como el TOEFL iBT y el GRE³, el proceso de evaluación es de manera mixta entre jueces humanos y electrónicos (mejor conocidos como *e-raters*). Otra manera de garantizar la confiabilidad es mediante evaluaciones simuladas en las cuales a todos los evaluadores se les da a corregir un texto que ya se encuentra calibrado, sin que ellos lo sepan, con el fin de establecer quienes son más severos o indulgentes al momento de emplear la rúbrica. En otros casos más, como los exámenes de alemán TestDaF y los de español con fines generales Diploma de Español como Lengua Extranjera (DELE), las correcciones se hacen por dos correctores humanos. Sin embargo, la problemática que se suscita al momento de evaluar los textos es la siguiente: ¿Qué sucedería si los dos evaluadores que corrigen un mismo texto fueran demasiado severos o indulgentes? El resultado que aportarían sería muy similar, y sus correlaciones altas; sin embargo, no reflejaría de manera precisa el desempeño de un sustentante. Por esta razón, se deben emplear herramientas estadísticas (como el análisis MFRM) que den cuenta de las posibles facetas o variables que inciden en aquellas pruebas en las cuales se evalúa el desempeño del candidato.

En caso de que haya discrepancia al momento de evaluar, es frecuente que un evaluador experimentado intervenga. En esta situación, se toma el puntaje más cercano al del corrector con mayor experiencia. En el caso de las rúbricas holísticas, se ha tomado como referente de discrepancia que exista más de un punto de diferencia entre los evaluadores; es decir, que si en la escala del 0 al 5 el primer evaluador le otorga un 5 al candidato y el segundo evaluador un 3, entonces intervendría un tercer evaluador. Esto equivaldría a un rango de tolerancia de aproximadamente un 20%. Sin embargo, en el caso de las rúbricas analíticas, el criterio de punto de discrepancia no ha sido claramente definido, aunque se podría adoptar la misma postura que para las holísticas. Estos rangos de discrepancia en las pruebas holísticas son los que evitan contar con un dato más fino del desempeño del evaluado, pero como se expuso en el apartado

³ Consúltese Breyer, Attali, Williamson, Ridolfi-McCulla, Ramineni, Duchnowski y Harris (2014) para información detallada de cómo se emplea el *e-rater* en el GRE. En el caso de este examen, el evaluador electrónico se emplea como una herramienta adicional después de la evaluación del primer humano. Si el evaluador electrónico detecta una diferencia mayor o menor a 0.5 puntos, entonces interviene un segundo evaluador humano. Entonces, el evaluador electrónico, por lo menos en este examen, se está empleando como instrumento de supervisión del trabajo humano.

anterior, la evaluación con estas escalas resulta pertinente cuando se tiene que evaluar a un gran número de sustentantes. En el caso de las rúbricas analíticas, el dato que se obtiene sobre el desempeño del candidato es más preciso, aunque como mencionan Jonsson y Svingby (2007), las rúbricas per se no facilitan el proceso de evaluación, pero sí pueden mejorar la confiabilidad cuando éstas son bien diseñadas, claras y detalladas, y los evaluadores cuentan con entrenamiento y constante monitoreo.

En suma, lo que deseamos al momento de evaluar es establecer una generalización de la evaluación; en otras palabras, que un sustentante obtendrá siempre un puntaje similar en una prueba específica siempre y cuando las tareas sean de dificultad similar, la evaluación por parte de los correctores sea también similar y las habilidades del candidato no se hayan modificado. Por ejemplo, si dicho candidato presentara nuevamente la misma prueba al día siguiente; es decir, sin una variación en sus habilidades de escritura, con otra versión y fuera evaluado por otro evaluador, su resultado debería ser muy semejante al del día anterior (consistencia externa o inter-evaluador). Más aún, si un mismo evaluador evaluara los mismos textos en tres ocasiones diferentes a lo largo de un año, sus resultados deberían ser también muy similares (consistencia interna o intra-evaluador). Entonces, lo que deseamos es que nuestros datos sean generalizables a estas variables.

5.2.2.1. LA CONSISTENCIA EN LOS RESULTADOS DE LOS EVALUADORES

Una de las mayores fuentes de invalidación en las tareas de respuesta construida se relaciona con la falta de consistencia en las evaluaciones. Hamp-Lyons (2007) menciona que el cuestionamiento sobre la confiabilidad de los evaluadores en PE comenzó en la década de los 80, pero que no fue sino hasta los 90 que se evidenció la necesidad de incrementar su confiabilidad. Consecuentemente, la confiabilidad en la evaluación de pruebas de respuesta construida mediante evaluadores es uno de los grandes retos que encara hoy en día la evaluación de la escritura. De hecho, Knoch (2007) asevera que los puntajes netos otorgados por un sólo evaluador son poco confiables; por lo cual, la autora aboga para que las evaluaciones sean hechas por al menos dos evaluadores o más, además de usar paquetes estadísticos como MFRM que permitan dar cuenta de la calidad del proceso de evaluación.

Para determinar el grado de confiabilidad de los evaluadores, tradicionalmente se han empleado diversas técnicas que buscan correlacionar los puntajes de los evaluadores. Sin embargo, Eckes (2009a) ejemplifica cómo estos métodos son erróneos, dado que parten del

supuesto que los evaluadores poseen el mismo grado de severidad, lo cual no es verdad. Este autor realizó un estudio con los evaluadores del examen de alemán con fines académicos, TestDaF, y calculó los índices de consenso entre los evaluadores (puntuaciones exactas y coeficiente Kappa de Cohen) y también el índice de consistencia (r de Pearson y Tau-b de Kendall⁴). El autor revela que los valores se encuentran muy por debajo del nivel satisfactorio; no obstante, lo que el autor desea destacar es que estos valores resultan irrelevantes, dado que se parte del supuesto mismo índice de severidad. Entonces, aquellos evaluadores que obtuvieron altas correlaciones podrían haber sido dos evaluadores demasiado severos, demasiado indulgentes o encontrarse en el centro, mientras que los evaluadores que presentaron bajos niveles de correlación podrían bien representar una gama de severidad distinta y, por ende, sus correlaciones son bajas. El autor afirma, entonces, que este problema genera una paradoja en la exactitud del consenso de los evaluadores:

High consensus or agreement among rater, and in this sense, high reliability, does not necessarily imply high accuracy in assessing examinee proficiency. Neither does high consistency imply high accuracy, even if consensus is high. Thus, high reliability may lead to the wrong conclusion that raters provide highly accurate rating when in fact they did not (Eckes, 2009a, p. 8).

Por las razones antes expuestas, Eckes (2009a) aboga por un enfoque de medición en vez del enfoque estándar de correlaciones entre evaluadores. El estadístico de facetas⁵ múltiples, conocido como MFRM, ha sido adoptado para resolver la paradoja que Eckes (2009a) plantea. El análisis MFRM se sustenta en un modelo conceptual que influencia la evaluación de los candidatos. La Figura 5.1 muestra la relación que se establece entre el texto producido por el examinado, la dificultad de la tarea y de las categorías de la rúbrica, y la severidad del evaluador. En la parte superior del modelo se ubica el factor más importante: la habilidad del sustentante denotada por el constructo. Del lado izquierdo se encuentran los factores distales (i.e., las características propias del examinado, el perfil del evaluador y las características de la administración del examen). Éstas variables no afectan directamente al proceso de evaluación, pero los factores pueden interferir entre sí y además también con los factores próximos, por lo cual afectan al proceso de evaluación indirectamente. En la parte central se ubican los factores próximos; es decir, aquellas variables o facetas de variabilidad presentes durante el proceso de

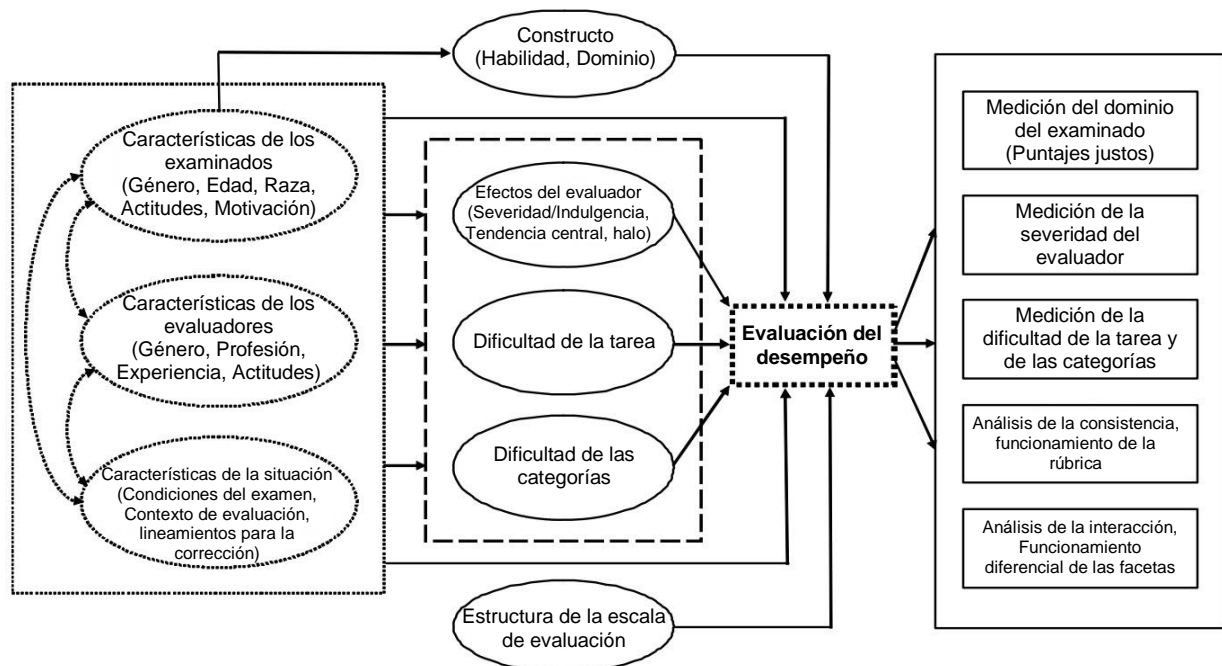
⁴ Consúltese Eckes (2009a) para mayor información detallada sobre el estudio que el autor llevó a cabo con la revisión de la prueba escrita del examen de alemán con fines académicos TestDaF.

⁵ Eckes (2009a) describe una faceta “como cualquier factor, variable o componente de una situación de medida que se asume puede afectar los resultados de un examen de una forma sistemática” (p.2).

evaluación y que pueden contribuir potencialmente al error de medida en la evaluación (i.e., los efectos del evaluador –severidad, efecto halo y centralidad–, la dificultad de la tarea y de las categorías de la rúbrica). El autor ubica en la parte inferior otro factor que resulta menos obvio y que se refiere a la estructura de la rúbrica empleada, misma que debería haber sido concebida con base en la definición del constructo. Sin embargo, esta variable no es de menor importancia y podría afectar negativamente la manera mediante la cual se la emplea para evaluar el desempeño de los candidatos. Finalmente, en la columna de la derecha se pueden apreciar los productos que se obtienen mediante el análisis de MFRM para PE: el resultado justo para cada uno de los evaluados, la severidad del evaluador, la dificultad de la tarea, la dificultad de las categorías de la rúbrica y el análisis de interacción entre las facetas.

Figura 5.1: Marco conceptual y psicométrico de los factores relevantes en la evaluación del desempeño mediada por los evaluadores (Eckes, 2009a, p. 11)

(Traducido por el investigador)



La dificultad de lograr una consistencia en la evaluación puede deberse a diversas causas: falta de claridad en los criterios descritos en la rúbrica, falta de pertinencia al incluir criterios que no pueden ser observados en el desempeño de los candidatos, falta de entrenamiento y capacitación para los evaluadores y las variables intrínsecas al evaluador (formación académica, estilo cognitivo, lengua materna, etc.). En cuanto a esta última variable, Eckes (2009a) menciona que el término variabilidad del evaluador “se refiere generalmente a la variación que se encuentra

asociada a las características del evaluador y no al desempeño de los examinados” (p. 4). De hecho, el autor matiza que las cuestiones asociadas a esta variabilidad son: los efectos del evaluador, los errores del evaluador y los sesgos del evaluador. Dentro de la primera categoría, el autor comenta que los efectos más notorios se deben a la severidad, el efecto halo y a la tendencia central. De estos tres, los últimos dos pueden deberse a que la rúbrica no es clara o pertinente para evaluar ciertos aspectos de la escritura, mientras que el primero hace referencia directa a las variables intrínsecas del evaluador. En cuanto a los errores del evaluador, Knoch (2007b) y Myford y Wolfe (2003) se refieren a éstos como inconsistencias o aleatoriedad en el uso de la rúbrica. Sin embargo, éstos podrían ser mitigados mediante la capacitación y el monitoreo de los evaluadores. En relación con los sesgos que se presentan en el evaluador, Knoch (2007) menciona que hay evaluadores que corrigen de forma muy severa ciertas categorías de la rúbrica, mismos que generalmente son los que consideran como centrales dentro de la escritura, mientras que otros que son concebidos como periféricos, son evaluados con menos dureza.

Wind y Engelhard (2013, p. 280) mencionan que los métodos empleados para garantizar la calidad de los evaluadores se pueden clasificar en tres categorías: la consistencia entre los evaluadores, los sesgos y errores sistemáticos del evaluador (su patrón de comportamiento) y la precisión del evaluador (la comparación entre el evaluador con aquel que es más experimentado). Más a fondo, dentro de la consistencia entre los evaluadores se pueden observar dos casos diferentes: inconsistencias entre evaluadores (inter-evaluador) e inconsistencias entre el mismo corrector (intra-evaluador). En el primer caso, los evaluadores otorgan diferentes puntajes a un mismo texto, mientras que en el segundo caso un mismo evaluador evalúa el mismo texto de forma diferente en dos ocasiones distintas, o evalúa textos con características semejantes de diferente forma.

Sin embargo, como apuntan los autores, hay ciertos efectos que no se pueden evidenciar en las comparaciones antes mencionadas; a saber: efecto halo o de tendencia central. Esto quiere decir que si un evaluador presenta efectos de tendencia central, puede correlacionar positivamente con otros evaluadores. Por ello, análisis estadísticos como MFRM proveen una valiosa herramienta que permite evidenciar otros efectos que pudieran presentar los evaluadores. En el estudio de Wind y Engelhard (2013, p.280), éstos evidencian que el evaluador más experto, y que se encuentra en el punto intermedio de severidad entre los evaluadores, también fue el que presentó los efectos de tendencia central más altos.

Eckes (2009a) menciona que hay que matizar dos conceptos de la consistencia inter-evaluador. El primero se refiere a la consistencia; es decir, al grado en el que los evaluadores le otorgan el mismo puntaje a un examinado, mientras que el segundo hace alusión a la confiabilidad, o al grado en que los evaluadores jerarquizan a los examinados de la misma manera. Según este criterio, dos evaluadores podrían ser confiables si jerarquizaran a los estudiantes de la misma manera, aunque con ligeras variaciones en sus puntajes.

Weigle (2002), siguiendo la línea de White 1984, describe ciertas técnicas que deben ser implementadas para evitar sesgos en los resultados obtenidos:

- Contar con una rúbrica que contenga explícita y claramente los criterios a evaluar.
- Ejemplos de desempeño que le sirvan al evaluador para establecer los diferentes niveles de actuación de un examinando.
- Idealmente, las evaluaciones deberían ser de forma controlada; es decir, que los evaluadores se reúnan para discutir posibles dificultades al momento de emplear la rúbrica o al evaluar casos atípicos de desempeño. No obstante, es raro que todas las sesiones tengan lugar de dicha manera, puesto que normalmente se conducen así al momento de la capacitación o en sesiones de calibración, pero posteriormente es difícil reunir siempre a los evaluadores para que corrijan y discutan al mismo tiempo los textos.
- Llevar un control de calidad de los evaluadores con el fin de darles seguimiento a quienes se encuentran fuera de los límites establecidos de confiabilidad o descartarlos en caso de ser necesario (p. 129).

5.2.2.2. LA CAPACITACIÓN DE LOS EVALUADORES

La relación que se establece entre la rúbrica, el evaluador y el texto en cuestión es indisoluble. Por esta razón, resulta primordial contar primeramente con un instrumento de medición (rúbrica) que evidencie de la manera más precisa posible lo que en realidad pretendemos evaluar. Si la rúbrica no es clara, es imprecisa o vaga en sus criterios, el evaluador –por muy experimentado que sea– encontrará dificultades para emplearla y basará su evaluación en criterios alternos no explicitados en la rúbrica (efecto halo, tendencia central o criterios holísticos). Entonces, si partiéramos del supuesto que la rúbrica ha sido diseñada adecuadamente, es clara y precisa en cada uno de sus rasgos, resta la interacción que surja entre el evaluador, el texto y el uso de la escala.

Las rúbricas representan un instrumento de medición; por lo tanto, el uso eficiente de la escala no se da por de facto; de hecho, la capacitación de los evaluadores resulta una tarea esencial con el fin de establecer acuerdos comunes de la manera en la cual se empleen cada uno de los criterios de la rúbrica para evaluar (Barkaoui, 2007; Becker, 2011; Eckes, 2009a; Huang y Foote, 2010; Knoch, 2007, 2009; Myford y Wolfe, 2003; Weigle, 2002). Knoch (2007b) comenta que los evaluadores entrenados ponen énfasis en los diversos componentes de la rúbrica de forma independiente, con lo cual se minimizan los juicios personales. East (2009) afirma que la demanda por contar con métodos válidos y confiables para evaluar la escritura en segundas lenguas ha crecido notoriamente en los últimos años, aunque uno de los grandes problemas que subyace a la falta de confiabilidad en los resultados obtenidos al momento de emplear una rúbrica, sin importar si ésta es holística o analítica, se debe a la falta de capacitación. Lo anterior no solamente pone en jaque la validez, sino que además atenta contra esta forma de evaluación como mecanismo para garantizar resultados más consistentes (Rezaei y Lovorn, 2010). Hamp-Lyons (2007) menciona que si bien es cierto que dicho entrenamiento es indispensable, los factores personales del evaluador pueden causar diferentes percepciones en éste, que inciden en la manera de observar y de evaluar un texto. Es indudable que los evaluadores no son máquinas, razón por la cual no se pueden eliminar por completo las diferencias entre éstos; sin embargo, la capacitación ayuda a los evaluadores a ser más conscientes y a que lleguen a acuerdos comunes sobre la implementación de los distintos criterios –con lo cual se puedan obtener resultados más consistentes–. De hecho, Bejar (2012) menciona que es indispensable que los evaluadores comprendan claramente la rúbrica y que la acepten, dado que aun cuando los criterios de la rúbrica son lo suficientemente claros y adecuados al propósito de evaluación, puede ser difícil para los evaluadores implementar la rúbrica al momento de evaluar.

En cuanto al tiempo que requiere un evaluador novato para convertirse en uno experimentado, Lim (2011) establece ciertas interrogantes que resultan pertinentes: “¿En qué momento están listos los candidatos para actuar como evaluadores? ¿Cuánto tiempo pueden fungir los evaluadores como tales sin necesidad de capacitación adicional, intervención alguna o apoyo? ¿Los efectos de la capacitación perduran? ¿Qué diferencia a un evaluador de aquellas personas que no lo son? ¿Cómo se convierte una persona en un evaluador?” (p. 545). Todas estas preguntas resultan imprescindibles al momento de decidir que una persona que ha recibido capacitación se encuentra lista para desempeñarse como un evaluador independiente.

5.2.2.2.1. LAS SESIONES DE CAPACITACIÓN

Normalmente, las sesiones de capacitación se organizan cuando hay nuevos evaluadores, cuando ha pasado un tiempo considerable entre el proceso de aplicación-corrección de un examen o cuando se han llevado a cambios o modificaciones sustanciales a la rúbrica. En cualquier caso, este procedimiento se realiza habitualmente mediante la evaluación de textos pertenecientes a aplicaciones anteriores o pilotajes en los cuales se han identificado previamente textos que son prototípicos de los distintos niveles de desempeño expresados en la rúbrica, o textos que son problemáticos o controversiales (DiPardo, Storms y Selland, 2011; Esfandiari y Myford, 2013; Jonsson y Svingby, 2007; Knoch, 2007b, 2009; Weigle, 2002). Toda vez que se ha concluido la sesión, se les comisiona a los evaluadores o futuros evaluadores otro tanto igual de textos para que los corrijan de forma independiente. Después de que la evaluación se ha hecho de forma individual, se reúne nuevamente el grupo para discutir los resultados obtenidos y limar diferencias o dudas que hayan surgido al momento de emplear la rúbrica. Posteriormente, dependiendo del rango de aceptación que hayan tenido los candidatos o evaluadores, se les proporcionan textos reales a evaluar (Esfandiari y Myford, 2013). Sin embargo, en caso de que haya evaluadores que sean inconsistentes en sus evaluaciones, deberán ser remplazados por otros. En caso de que sean necesarias sesiones posteriores de calibración, éstas se realizan para lograr la mayor consistencia posible en los resultados de los correctores. Finalmente, Weigle (2002) sugiere que se elabore una bitácora en la cual se puedan registrar algunos de los problemas particulares surgidos al momento de evaluar y en donde se anoten las decisiones que se adoptaron para resolver dicho caso. Por ejemplo, en el caso de que un examinando escriba con letra ininteligible, o que su texto sea adecuado pero incompleto o que simplemente se aborde una temática completamente ajena a la solicitada.

Pese a que la capacitación de los evaluadores resulta esencial (Eckes, 2009a; Hamp-Lyons, 2007; Knoch, 2007; Rezaei, y Lovorn, 2010; Weigle, 2002), no necesariamente ésta garantiza mayor consistencia y confiabilidad en las evaluaciones (Attali et al., 2012; Eckes, 2009a, 2012; Prieto, 2011). De hecho, Knoch (2007) menciona que después de haber recibido capacitación, sólo un pequeño número de evaluadores demostró menor sesgo en sus resultados, aunque otros desarrollaron otros sesgos nuevos. A este respecto, Rezaei y Lovorn (2010) afirman que cuando no hay capacitación para los evaluadores, el uso de la rúbrica se convierte tan sólo en una lista de comprobación (*checklist*). Huang y Foote (2010) también identificaron que una fuente potencial de variabilidad en los resultados obtenidos se debía a la falta de entrenamiento de los evaluadores en el uso de la rúbrica; sobre todo, dado que los evaluadores

provenían de antecedentes profesionales muy diversos. Sin embargo, las sesiones de capacitación son demasiado cortas, y en el mejor de los casos duran unas cuantas horas (Lim, 2011).

En realidad, el fin último de la capacitación consiste en que los correctores se familiaricen con el constructo para el cual se desea evaluar y lleguen a un entendimiento mutuo sobre qué muestras o ejemplos corresponden a los niveles de desempeño descritos en cada una de las categorías de la rúbrica. Aunque también es cierto que si la rúbrica no es clara ni precisa en sus descripciones del nivel de actuación o logro del sustentante, difícilmente habrá mejoras en la confiabilidad de los correctores. De hecho, Eckes (2009a) menciona que existen numerosos casos en los cuales, pese a que se les ha dado entrenamiento exhaustivo y retroalimentación a los evaluadores en cómo usar la rúbrica, los niveles de confiabilidad no varían o incluso empeoran. Por añadidura, el autor comenta que incluso aquellos evaluadores más experimentados presentan variaciones en el uso de las rúbricas. En suma, pese a que la capacitación de los evaluadores puede mitigar hasta cierto punto las diferencias de severidad entre éstos, las discrepancias entre evaluadores nunca desaparecerán por completo. Por consiguiente, el valor último de la capacitación radica en que los evaluadores sean más conscientes de los criterios a evaluar y cómo evaluarlos.

En contraposición a lo que autores como Dipardo et al. (2011), Knoch (2007), Weigle (2002) y Zhao, (2012) comentan sobre la importancia en la capacitación de los evaluadores, Attali et al. (2012) aseveran que en comparación con el tiempo que le toma a un sustentante escribir un ensayo (típicamente 30 minutos), la confiabilidad por parte de los evaluadores es mediocre; razón por la cual se ha optado por utilizar correctores electrónicos (*e-raters*) en el caso de exámenes administrados a gran escala como el *Graduate Management Admission Test* (GMAT), GRE y TOEFL iBT. Sin embargo, aun cuando los *e-raters* son confiables, también presentan deficiencias al no poder evaluar aspectos superiores de escritura que un humano si puede evaluar⁶. El foco de la presente investigación no pretende ahondar en los *e-raters*⁷, aunque el estudio comparativo

⁶ Attali et al. (2012, p.127) mencionan que estos aspectos superiores de escritura se relacionan con la calidad de las ideas, su desarrollo y organización. Según estos autores, los humanos pueden evaluar dichos aspectos puesto que son capaces de comprender el texto, mientras que los *e-raters*, no.

⁷ Consúltese Lee et al. (2008) para mayor información sobre el proceso de comparación entre tres mecanismos diferentes de evaluación: mediante una rúbrica holística, una analítica y con un procesador automático (*e-rater*). Los autores encontraron altas correlaciones entre el uso de la rúbrica holística, la analítica y el *e-rater*.

de éstos con su contraparte humana (Attali et al., 2012) da cuenta de cómo podría complementarse la evaluación por parte de jueces humanos y por correctores electrónicos. En el estudio, los autores buscaron que los evaluadores enfocaran su atención en los aspectos superiores de escritura. Para ello, emplearon una rúbrica modificada que enfatizaba dichos aspectos en relación con la rúbrica original. Los investigadores encontraron que cuando los evaluadores ponían especial énfasis en los aspectos superiores de escritura, en el caso de ensayos argumentativos, la correlación inter-evaluadores era mayor. En conclusión, estos autores proponen un sistema dual que funcione más bien de mecanismo complementario; en otras palabras, que los correctores humanos pongan especial énfasis en aspectos superiores de escritura, mientras que su contraparte electrónica analice aspectos inferiores de escritura (tales como la ortografía, el léxico o la gramática). Breyer et al. (2014), por su parte, consideran que más bien los procesadores naturales del lenguaje pueden servir como complemento de aspectos relevantes considerados dentro del constructo del examen.

The advances in automated essay scoring (AES) and its operational implementation underscore the need for a "division of labor" between human and machine assessment that takes into account the relative strengths of both methods. By encouraging raters to focus on higher-order aspects of writing it might be possible to achieve (a) higher reliability of human scoring, (b) lower correlations between human and machine scores, and consequently (c) higher reliability of combined human and machine scores (Attali et al., 2012, p. 127).

De hecho, uno de los hallazgos más significativos de estos autores, en sintonía con los de Alderson et al. (1995), fue que el aumento de las categorías en la escala favoreció la evaluación de los correctores. Aunado a lo anterior, Attali et al. (2012) expresan que los evaluadores no recibieron ningún tipo de capacitación adicional con la nueva rúbrica y, sin embargo, la correlación en sus evaluaciones fue satisfactoria, con lo cual se comprueba que los evaluadores experimentados pueden corregir mejor los textos con una escala detallada y fina (véase Knoch, 2007, 2009, Weigle, 2002). Finalmente, los autores también sugieren que los ensayos sean organizados de acuerdo con su longitud, puesto que cuando el evaluador corrige ensayos con una longitud similar, éstos pueden discernir mejor aquellos aspectos superiores de escritura y, por lo tanto, otorgarle mejor una puntuación al texto. Por último, Eckes (2012) afirma que al conocer el estilo cognitivo del evaluador, se pueden plantear objetivos al momento de conducir talleres de entrenamiento, en aras de lograr mitigar las diferencias entre los evaluadores y así, lograr una percepción balanceada de la importancia de los criterios.

En suma, lo que se puede concluir de la capacitación de los evaluadores es que si bien ni ésta ni la retroalimentación sobre su desempeño fomentarán que los índices de severidad sean

iguales, sí se establecen puntos de común acuerdo que coadyuvan a la evaluación de los textos y a la consistencia de los evaluadores a corregir los textos de manera sistemática.

5.2.2.3. EL MONITOREO DE LOS EVALUADORES

Dado que la evaluación de los evaluadores presenta múltiples fuentes de variabilidad que inciden al momento de evaluar los textos, el constante monitoreo de éstos es indispensable. Bejar (2012) comenta que al momento de evaluar, el evaluador se forja una representación mental tras haber leído el texto. En ese momento, el evaluador coteja su representación mental del texto con la rúbrica y asigna un valor a cada uno de los criterios. Sin embargo, esta asignación depende de múltiples variables: la verdadera calidad de la respuesta, la calidad de la representación que se ha formado el evaluador, la información previa que el evaluador ha acumulado durante la sesión de evaluación, el estado físico y mental del evaluador, las condiciones ambientales en las cuales se está evaluando y el tipo de respuestas previamente evaluadas (p. 5). Por estas múltiples razones, es indispensable darles seguimiento continuo a los evaluadores, sin importar lo experimentados que sean o lo consistentes que hayan sido previamente con sus evaluaciones. De hecho, Lim (2011) destaca que si bien es cierto que los evaluadores experimentados demuestran un mejor desempeño que los novatos, gracias a la capacitación y a la práctica, éstos últimos pueden rápidamente lograr niveles de confiabilidad tan aceptables como los experimentados. Más aún, el autor afirma que la mejora en la consistencia de los evaluadores novatos podría estar relacionada con el volumen de textos que se evalúan y no necesariamente con la capacitación que éstos reciben. El investigador condujo un estudio con el examen de inglés *Michigan English Language Assessment Battery* (MELAB). Este examen certifica oficialmente a un evaluador de PE solamente cuando ha logrado demostrar consistencia consecutiva en 80 textos, en los cuales no hay discrepancia mayor a un 10%. Este proceso puede tardar incluso varios meses. Mediante un estudio longitudinal, se buscó identificar en qué momento un evaluador novato se hacía consistente en sus evaluaciones, comparado con uno experto. En un período de un año, se efectuaron mediciones en tres momentos diferentes. Lim (2011) identificó que los evaluadores novatos lograron consistencia después de 130 textos evaluados. El autor concluye que quizás una de las razones principales por las cuales los evaluadores novatos se vuelven consistentes se deba al volumen de textos por evaluar y también a la constancia.

Wind y Engelhard (2013) mencionan que un estudio cualitativo en aquellos evaluadores que se encuentran fuera de los parámetros aceptables de evaluación podría arrojar información

valiosa sobre cuál podría ser la causa de las discrepancias o de los índices de severidad o indulgencia de los correctores. Los autores, siguiendo la línea de Lumley (2002), resaltan que el evaluador, no la rúbrica, es quien se encuentra en el centro del proceso de evaluación, puesto que éste es quien decide a qué rasgos poner atención al momento de evaluar, cómo lidiar con conflictos inevitables de discrepancia entre la rúbrica y el texto, y cómo justificar los valores adjudicados a cada uno de los criterios de la rúbrica (p. 297).

5.2.2.4. EL PROCESO DE EVALUACIÓN DE TEXTOS

La interacción que se presenta entre el evaluador, el texto y la rúbrica son indisociables y también las variables que eventualmente entrarán en juego para la asignación de los puntajes. Sin embargo, pocos son los estudios que se han llevado a cabo para evidenciar el proceso de interpretación de los rasgos de la rúbrica y de selección de puntajes por parte del corrector. Hasta la fecha, uno de los estudios cualitativos más socorridos que se han empleado para comprender la forma en que los evaluadores realizan la corrección del texto es mediante el método para pensar en voz alta (*think-aloud protocol*) (Barkaoui, 2007, 2010a, 2010b, 2011; Knoch, 2007, 2009; Zhao, 2012). Bajo este mecanismo, el evaluador –directa o indirectamente (mediante una grabación)– describe los criterios que está tomando en cuenta al momento de evaluar el desempeño de un candidato en una prueba escrita; con lo cual el investigador puede seguir la ruta del proceso cognitivo que está empleando el evaluador y de las consideraciones que está adoptando para evaluar el texto.

Hamp-Lyons (2007, p. 6) recalca que poco se ha logrado identificar en cómo los evaluadores llegan a las decisiones que toman y mucho menos al establecimiento del consenso entre un grupo de evaluadores al momento de revisar textos de manera conjunta. De hecho, en el estudio que la autora condujo en sesiones entre los evaluadores para llegar a consensos sobre el puntaje óptimo, menciona que las grabaciones no proporcionaron una solución, sino más bien dieron una clara evidencia del problema para llegar a consensos. Incluso menciona que los evaluadores admitían salir de las sesiones con sentimientos de ansiedad, confusión e incluso frustración al no poder sentir una verdadera estandarización incluso cuando las sesiones concluían con el establecimiento de puntajes similares. Debido a esto, la autora aboga por una aproximación distinta a estas sesiones de emparejamiento y común acuerdo entre los evaluadores: un compartir positivo (*positive sharing*). A diferencia de un simple intercambio, este compartir positivo busca que los evaluadores expongan las cualidades del ensayo (fundamentalmente en aquellos controversiales) con miras a convencer a sus colegas sobre su

perspectiva. Cada una de estas perspectivas es considerada como distintas verdades y, por lo tanto, un puntaje justo se encuentra en algún punto de la interacción de estas verdades.

Dentro de los hallazgos más significativos en el uso de los pensamientos en voz alta, autores como Zhao (2012, p. 208) mencionan que, pese a que el corrector emplea una rúbrica analítica, lo primero que los evaluadores reportan sobre lo que observaron en el texto es si éste tenía sentido y si éstos podían seguir el discurso del sustentante. Aunado a lo anterior, el uso de ejemplos creativos y también la inclusión de detalles en textos argumentativos resultaba relevante para los evaluadores, de primera impresión. Estos hallazgos van de la mano con investigaciones previas (Cumming, Kantor y Powers, 2001) que permitieron identificar que además del énfasis en el contenido, la retórica y el lenguaje, los evaluadores también se auto-monitorean al releer el texto y al compararlo con otros.

Investigadores como Barkaoui (2007, 2010, 2011) han conducido diversos estudios que permitan evidenciar cómo es el proceso de evaluación de textos por parte de los correctores. Empleando protocolos en voz alta, Barkaoui (2010a) condujo un estudio para evidenciar los procesos de toma de decisión al momento de emplear una rúbrica para evaluar textos escritos en inglés como lengua extranjera. El investigador incluyó a 11 evaluadores novatos y 14 experimentados, quienes mediante rúbricas holísticas y analíticas corrigieron, cada uno, 12 ensayos en silencio y 12 en voz alta. En una primera fase, los evaluadores emplearon la rúbrica holística y dos semanas después, la analítica para evaluar los mismos textos. El autor destaca que más que la experiencia, pareciera que el tipo de rúbrica genera un proceso cognitivo diferente. Al evaluar con una rúbrica holística, los jueces ponen mayor atención al texto y recurren a éste para interpretar aquellas secciones que nos son claras. Sin embargo, en el caso de las rúbricas analíticas, los evaluadores recurren mayoritariamente a la escala para asignar el puntaje de cada uno de los criterios y justificar la asignación de los niveles. Debido a que la rúbrica analítica contiene diversas categorías, el autor menciona que los evaluadores son más consistentes en sus evaluaciones, pues tienen que poner atención a cada uno de los aspectos que se evalúan en los textos. En cuanto a la diferencia entre evaluadores novatos y experimentados, uno de los hallazgos del investigador fue que los novatos eran más dependientes a la rúbrica que los experimentados. No obstante, además de que ambos grupos de evaluadores reportaron una denotada preferencia hacia la rúbrica analítica (véase también Knoch, 2009), este tipo de rúbricas permite que los evaluadores reduzcan los conflictos al momento de asignar valores holísticos cuando los criterios de la escala entran en conflicto

(Barkaoui, 2010a, p. 68). Consecuentemente, resultan más adecuadas para evaluadores novatos. Otra diferencia sustancial fue que los evaluadores novatos eran menos predecibles y más propensos a enfocarse en el contenido, mientras que los evaluadores más experimentados daban mayor importancia a la precisión lingüística y eran más severos (Barkaoui 2010b).

Sin embargo, en otro estudio sobre el proceso de corrección del texto, Barkaoui (2011) menciona que una tercera parte de los evaluadores experimentados encontraron problemas para verbalizar sus pensamientos durante los protocolos de pensamiento en voz alta. Algunos de los argumentos expuestos por los correctores fueron que muchos de estos pensamientos se encontraban automatizados y, por lo tanto, resultaba difícil verbalizarlos. Adicionalmente, la sensación de estar siendo grabados también inhibía a los evaluadores o limitaba que éstos se expresaran libremente con respecto a lo que estaban pensando. Asimismo, algunos de los evaluadores manifestaron perder la concentración al leer en voz alta y expresar sus pensamientos también en voz alta; por lo cual debían releer varias veces el texto. Por último, el autor también menciona que los índices de severidad varían cuando se evalúan los textos mediante protocolos de pensamiento en voz alta. Es decir, algunos evaluadores se mostraron más severos, mientras que otros menos. El autor concluye que este tipo de estudios cualitativos afecta la severidad y consistencia de algunos de los evaluadores, por lo que quizás debería optarse por otros mecanismos como la entrevista rememorada (*stimulated recall*). En este proceso, el evaluador es entrevistado justo después de haber corregido el texto para reportar de qué forma lo evaluó y bajo qué criterios.

En otro sentido, DiPardo et al. (2011) destacan que pese a que las correlaciones entre los evaluadores fueron altas, en las entrevistas y observaciones a las cuales fueron sujetos, los evaluadores incorporaron en sus evaluaciones rasgos que no se encontraban descritos en las rúbricas. Este hecho pone especial énfasis en que aunque los evaluadores llegaron a resultados similares, no se debió necesariamente a un entendimiento compartido de las categorías de la rúbrica o del constructo de escritura subyacente a la prueba. Por esta razón, se requieren más estudios que den cuenta de cómo interpreta la escala el evaluador al momento de leer un texto y asignarle un puntaje.

5.2.2.4.1. EL EVALUADOR ELECTRÓNICO

En los ECLAA administrados a gran escala, cada vez es más común que se empleen evaluaciones de ensayos automatizadas, *Automated Essay Scoring*, (AES), conocidos también

como evaluadores electrónicos (*e-raters*). Este tipo de herramientas permiten evaluar una vasta cantidad de textos en tan solo unos minutos, por lo cual resultan prácticos y sumamente económicos de usar —en comparación con los evaluadores humanos—. Enright y Quinlan (2010, p. 319) mencionan que los *e-raters* son capaces de reconocer características como: la gramática, el uso, la mecánica, el estilo, la organización, el desarrollo, la complejidad léxica, el vocabulario pertinente para una consigna específica y el manejo de este vocabulario (contenido).

Hamp-Lyons (2014) menciona que actualmente existe una creciente disyuntiva entre el uso de *e-raters* y evaluadores humanos. Esto se debe a que los primeros son fáciles de implementar y también sistemáticos; sin embargo, la autora advierte que el remplazo de los humanos en la evaluación de textos, dejaría fuera el sentido de la escritura: escribir en un contexto social y para una audiencia específica. De hecho, uno de los argumentos más asiduos en contra de los evaluadores electrónicos es que la “escritura para ser leída por una máquina viola la naturaleza de la esencia social de la escritura” (Deane, 2013, p. 8). Este autor destaca el riesgo de emplear AES cuando los alumnos van a ingresar a la universidad, dado que si su primera experiencia con la escritura académica es ser leídos y evaluados por una máquina, esto puede fomentar una concepción errónea sobre la naturaleza de la escritura como un constructo social destinado a una cierta comunidad o audiencia.

De manera muy puntual, los AES o *e-raters* se emplean para evaluar ensayos de los exámenes del TOEFL iBT, el GMAT y el GRE. Éstos son quizás los exámenes de lengua con fines académicos en inglés que mayor demanda tienen a nivel mundial. En el caso del GRE y del TOEFL iBT, el *e-rater* ha tomado el lugar del segundo evaluador (Enright y Quinlan, 2010), y solamente cuando existe discrepancia entre la máquina y el humano, interviene un segundo evaluador humano. En caso de que el primer evaluador humano identifique que el texto del examinado se encuentra fuera del tema solicitado, entonces en vez de que éste sea evaluado por el *e-rater*, es evaluado por un segundo juez humano. Cabe mencionar que en otras lenguas como el chino, el francés, el alemán y el español se está también trabajando en la elaboración de AES (Deane, 2013). Sin embargo, las controversias que se han suscitado en torno a la implementación de éstos resultan relevantes. Weigle (2010) menciona que, sobre todo, el temor del uso e implementación de los *e-raters* es por parte de los docentes que imparten cursos de escritura y también por los estudiantes. Por una parte, se arguye que los evaluadores humanos son inconsistentes, presentan correlaciones bajas y sesgos debidos a múltiples causas (p.ej., lengua materna, origen geográfico, antecedentes académicos, experiencia, edad, entre otras),

mientras que los evaluadores electrónicos son altamente confiables en los resultados que arrojan (Attali et al., 2012). En contraparte, existe una gran polémica sobre la naturaleza misma de la escritura y de la visión de ésta como una construcción social encaminada a una determinada audiencia humana que dista mucho de las máquinas que se emplean para evaluar los ensayos.

Deane (2013) presenta una interesante discusión sobre los evaluadores electrónicos y su pertinencia dentro del proceso de evaluación. Por una parte, comenta el autor, existen fuertes oposiciones a la inclusión de AES, sobre todo ante el temor de que éstos eventualmente reemplacen a los humanos, pues éstos primeros no vislumbran la escritura como una forma de comunicación. Deane (2013) y Weigle (2010) añaden que más que abogar a favor o en contra de éstos y de compararlos con sus contrapartes humanas, es importante establecer un justo balance en lo que pueden aportar a la corrección de textos. Por ello, más que investigar las correlaciones entre los evaluadores electrónicos y los humanos, debería establecerse un trabajo colaborativo entre ambos, pues es cierto que las máquinas pueden evaluar de forma más eficaz y computarizar de manera muy precisa ciertos aspectos sintácticos, léxicos y morfológicos de manera automática y certera, mientras que los humanos pueden enfocarse a aquellos aspectos que se relacionan con el pensamiento crítico, la retórica textual, el contenido de las ideas y de la creatividad empleada al momento de leer un texto (Enright y Quinlan, 2010). De hecho, en el estudio de Weigle (2010) para establecer la comparación entre los evaluadores humanos y su contraparte electrónica, la autora menciona que aun cuando las correlaciones entre ambos son adecuadas, el *e-rater* no puede duplicar las evaluaciones humanas. McCurry (2010), sin embargo, encontró que los resultados obtenidos por los evaluadores electrónicos diferían significativamente de los humanos en la evaluación de ensayos abiertos como los que proponen el TOEFL o IELTS. De hecho, el autor concluye que el *e-rater* no puede evaluar de manera tan confiable como un humano.

Uno de los problemas potenciales de los AES es que cuando se trata de exámenes a gran escala, y por lo tanto de alto impacto, los candidatos podrían identificar al evaluador (en este caso la máquina) más bien como una barrera que esquivar en vez de una persona (evaluador humano) a la cual comunicar algo. Esto conlleva importantes implicaciones dado que los examinados podrían emplear ciertas estrategias para “burlar a las máquinas”; por ejemplo, empleando vocabulario sofisticado, pero inapropiado para el texto, o frases agramaticales, pero de alta complejidad sintáctica con el fin de obtener un mejor puntaje. Esto, sin embargo, es percibido por un evaluador humano que, aunque no es capaz de identificar con certitud el número

de palabras académicas o con sofisticación léxica que ha empleado un candidato, si puede determinar la pertinencia del uso de ciertas palabras en un contexto específico y con un fin comunicativo claro. Deane (2013) recalca que la escritura no es un proceso lineal, sino más bien un proceso sociocognitivo en el cual se ven involucrados otros procesos como la reflexión, la metacognición y el uso de estrategias que se emplean para transformar el conocimiento (p. 16).

En resumen, la evaluación de alto impacto administrada a gran escala está revolucionando la forma mediante la cual se evalúa. Si a esto le agregamos que los evaluadores humanos despliegan ciertos efectos perniciosos al evaluar (i.e., efecto halo, centralidad e índices de severidad disímiles), la presión por evaluar los textos producidos en ECLAA eficientemente, con precisión y consistencia, ha conducido a la incorporación de los *e-raters* dentro del proceso de evaluación. Éstos representan una valiosa herramienta y son de gran utilidad, además de que su avance tecnológico es exponencial, no obstante, su uso debe ser cuidadoso y meditado, además de encontrarse siempre bajo la supervisión de los humanos.

5.2.2.5. VARIABLES QUE INFLUYEN EN LA CONSISTENCIA DE LOS EVALUADORES

El evaluador juega un papel central dentro del proceso de asignación de un puntaje; sin embargo, los factores que contribuyen a la variabilidad en los evaluadores son numerosos. Estas variables pueden ser extrínsecas o intrínsecas. En cuanto a las primeras, éstas dependen del momento de la evaluación, y generalmente están relacionadas con la fatiga producida al evaluar ensayos a gran escala (Bejar, 2012). En estas circunstancias, debido a la confidencialidad de la información, es común que se congregue a un amplio grupo de correctores en una misma sala para evaluar durante una o dos semanas consecutivas (incluso sábados y domingos). Consecuentemente, los evaluadores pueden presentar alteraciones en sus evaluaciones al principio y al final del día o a lo largo de los días. En el caso de las variables intrínsecas; por ejemplo, estas tienen que ver con el evaluador en sí: su experiencia profesional, estilo cognitivo, rasgos personales, características demográficas, carga de trabajo, lengua materna y las estrategias que emplea para asignar los puntajes de la rúbrica (Bejar, 2012; Eckes, 2009a; Huang y Foote, 2010; Lim, 2011; Myford, 2012; Wolfe y McVay, 2012;). A este respecto, Lumley (2005) menciona que los evaluadores emplean su conocimiento e intuición para resolver discrepancias cuando los textos contienen ciertos elementos que no han sido descritos por las rúbricas. En efecto, como menciona Knoch (2009), las rúbricas son ambiguas y la evaluación resulta impresionista, por lo cual se requieren más estudios sobre cómo aspectos superficiales afectan a los evaluadores. La pregunta entonces es: ¿Qué estrategias emplea un corrector para evaluar

a los examinados? Esto no quiere decir que ni la capacitación ni la rúbrica estén de más, pero sí hay numerosas características intrínsecas y extrínsecas al evaluador que demarcan la forma en la cual percibe y corrige un texto. Estas variables se encuentran aún poco exploradas, aunque las que recientemente se han estudiado corresponden a los antecedentes profesionales del evaluador y su estilo cognitivo.

5.2.2.5.1. LOS ANTECEDENTES DEL EVALUADOR

Huang y Foote (2010) comentan que el antecedente académico y laboral incide particularmente en la manera mediante la cual los evaluadores corrigen con mayor o menor severidad las diversas categorías de las rúbricas. Asimismo, mencionan que la experiencia también incide en cómo los textos son percibidos por los evaluadores, afectando sus índices de severidad a lo largo del tiempo. De hecho, estos autores afirman que hacen falta estudios que den cuenta de los factores personales que inciden en el nivel de severidad al momento de evaluar y también en la estabilidad de los evaluadores a lo largo del tiempo. Eckes (2012) identificó una relación directa entre la importancia que le otorga un evaluador a una categoría de la rúbrica analítica, dado que éste considera que es importante para evaluar la calidad del texto, y la severidad con la que la evalúa. Entonces, según este autor, cuando los evaluadores son más severos, se debe a que éstos consideran que este aspecto de la rúbrica es crucial para evidenciar las habilidades de escritura de los sustentantes. Por su parte, He, Gou, Chien y Chen (2013) arguyen que uno de estos factores de variabilidad en la percepción de las categorías afecta la implementación de la rúbrica. Por consiguiente, su severidad o indulgencia en la evaluación de éstas se encuentra determinada por la formación académica de los evaluadores. Estos autores mencionan que la procedencia de diversas disciplinas académicas genera diversas idiosincrasias y valores jerárquicos distintos en la manera en que se perciben las categorías de una rúbrica analítica. Aun cuando los autores encontraron diferencias en evaluadores procedentes de cuatro disciplinas académicas (lingüística, traducción, enseñanza de inglés como lengua extranjera y literatura), este estudio se encuentra limitado por la muestra. En realidad, la teoría de que la formación académica incide en cómo se percibe y evalúa cada una de las categorías de la rúbrica es interesante; no obstante, sería importante contar con una muestra mucho más amplia de evaluadores pertenecientes a este grupo de disciplinas académicas.

Estas fuentes de variabilidad intrínsecas al evaluador interactúan con el uso de la rúbrica y generan efectos que repercuten en mayor o menor medida la consistencia de los evaluadores. Entre éstos destacan: el efecto halo, el efecto de tendencia central, la incongruencia y la

severidad del evaluador. Bien a bien no se ha investigado la interacción entre los antecedentes del evaluador y los efectos que producen, aunque éstos se han tratado de sistematizar. El estudio de Wolfe y McVay (2012) consistió en evidenciar hasta qué punto un grupo de evaluadores presentaban efectos que potencialmente serían causantes de sesgos en los puntajes otorgados a los examinados. De hecho, con el fin de mostrar de manera práctica cómo se verían estos efectos al momento de emplear una rúbrica, los autores proporcionan los ejemplos que se presentan en el Cuadro 5.2. En éste se puede observar claramente cómo se verían los comportamientos de los evaluadores en tres efectos distintos: severidad/indulgencia del evaluador, tendencia central e incongruencia en la corrección. De estos tres, los dos primeros hasta cierto punto son aceptables y se pueden corregir con programas estadísticos y la retroalimentación pertinente, mientras que el tercero indica que por alguna razón el evaluador no es apto para evaluar los textos. Como se puede apreciar en el Cuadro 5.2, el ejemplo 1 está dado por un evaluador que es más indulgente que el promedio. En el contraste entre ambos datos se puede observar cómo los resultados correlacionan adecuadamente con el promedio, pero son ligeramente más altos. En el segundo caso se observa nuevamente un patrón de ascendencia de cada uno de los criterios, pero se observa como el evaluador tiende a excluir las puntuaciones que son muy altas o muy bajas (0 y 4) y permanece en las centrales (efecto de tendencia central). Finalmente, en el tercer caso, se aprecia como el evaluador es completamente asistemático en la manera en la que otorga sus calificaciones en comparación con lo que se consideraría adecuado.

Cuadro 5.2: Efectos de los evaluadores (Wolfe y McVay, 2012, p. 32)

(Traducido por el investigador)

1	Evaluaciones precisas Evaluaciones indulgentes	0 0 0 1 1 1 2 2 2 3 3 3 4 4 4 0 0 1 1 2 2 2 3 3 3 4 4 4 4 4
2	Evaluaciones precisas Evaluaciones de tendencia central	0 0 0 1 1 1 2 2 2 3 3 3 4 4 4 0 1 1 1 2 2 2 2 2 2 2 3 3 3 4
3	Evaluaciones precisas Evaluaciones imprecisas	0 0 0 1 1 1 2 2 2 3 3 3 4 4 4 2 2 3 1 4 4 3 2 3 1 4 1 1 4 2

Para poder evidenciar los efectos del evaluador, los autores proponen una serie de análisis estadísticos que permitan dar cuenta independiente de cada uno de estos efectos. De manera muy puntual, los autores comentan que el análisis estadístico MFRM en FACETS (Linacre, 1989) permite dar cuenta de estos efectos. En el caso de la calidad de los evaluadores, el rango convencional establecido (Linacre, 2012b) es de 0.5 a 1.5, aunque Linacre matiza que para los

exámenes de alto impacto, el rango debería oscilar entre 0.8 y 1.2. La tendencia central se puede obtener mediante la correlación de los residuales esperados ($r_{exp,res}$) cuyo valor preestablecido (Linacre, 2012b; Wolfe y McVay, 2012) es significativo cuando éste es menor a -0.3. En el caso de la incongruencia en los resultados otorgados por los evaluadores, éstos pueden identificarse cuando los valores de ($r_{exp,res}$) son positivos. Con el fin de identificar problemas de sesgo en ciertos criterios de la rúbrica o debido a causas superfluas en la evaluación, los autores proponen que se emplee el PCA en la matriz de los residuales. Las cargas se consideran como significativas cuando presentan valores mayores a 0.4.

5.2.2.5.2. LA COGNICIÓN DEL EVALUADOR

En los últimos años, uno de los aspectos que más ha llamado la atención de los investigadores en cuanto a las variables del evaluador se relaciona con la cognición de éstos. Tanto estudios cualitativos –mediante cuestionarios y protocolos de pensamiento en voz alta (*think-aloud protocols*)– como cuantitativos (p.ej., MFRM o la Teoría de la Generalizabilidad) han tratado de aportar información sobre cómo los evaluadores perciben el texto. Sin embargo, la relación entre esta información y su comportamiento al momento de evaluar textos es escasa.

De los estudios cualitativos, uno de los rasgos que ha resultado más sobresaliente es el estilo del evaluador. Ya McNamara (1996) desde hace casi veinte años había señalado que quizás la diferencia en los evaluadores se debía a que existían correctores con estilos cognitivos diferentes. Eckes (2009a, 2012), con miras a investigar un fenómeno de los evaluadores poco estudiado: la relación entre la cognición y el comportamiento del evaluador al corregir un texto, retoma el modelo de Wolfe (1997) para comprender cómo se establece dicha relación. El modelo que Wolfe (1997, p. 89) propone está conformado de dos componentes distintos pero interrelacionados entre sí: el marco de la evaluación y el marco de la escritura. En el marco de la evaluación, el corrector se hace una representación mental del texto mientras éste es leído. En el segundo marco, el evaluador se hace una representación mental de las categorías contenidas en la rúbrica. Al momento de evaluar, se establece una interrelación entre estos componentes. Aquí es donde según el modelo de Wolfe (1997) subyacen las diferencias en la asignación de los puntajes, ya que las representaciones mentales tanto del texto como de las categorías de la rúbrica varían entre un evaluador y otro. Aunado a esto, Eckes (2012) señala que el proceso de los evaluadores más experimentados es de arriba hacia abajo (*top-down*), poniendo atención primariamente al desempeño, mientras que el de aquellos menos experimentados es de abajo hacia arriba (*bottom-up*), poniendo más atención en aquellos rasgos más específicos.

En cuanto a los estudios cualitativos, los protocolos de pensamiento en voz alta han sido usados para investigar el aspecto cognitivo de los evaluadores en el proceso de corrección de los evaluadores (Myford y Wolfe, 2003, Wolfe, 2004). Sin embargo, en los últimos años se han empleado cuestionarios para evidenciar cómo es este estilo. En este sentido, Eckes (2008b) realizó un estudio con 64 evaluadores del examen de alemán con fines académicos. El investigador les pidió a los evaluadores que puntuaran del 0 al 4, en nivel de importancia, las distintas categorías de la rúbrica analítica empleada para corregir los textos. De este estudio, el autor identificó seis tipos de evaluadores; a saber: el de tipo sintáctico, de corrección, de estructura, de fluencia, de no fluencia y de no argumentación. Esto denota que la cognición de evaluador impacta en la forma en la que evaluarán los textos.

En otros estudios mixtos, He et al. (2013) utilizaron un cuestionario para identificar la importancia que le dan los evaluadores a cada una de las categorías de una rúbrica analítica. Los autores emplearon una escala tipo Likert del 0 al 10 (cero para el rango menor y diez para el mayor) con el objetivo de que identificaran la importancia de cada una de las categorías de la rúbrica. Comparando el análisis cuantitativo mediante MFRM y la escala tipo Likert, los autores identificaron que las categorías que los evaluadores puntuaron en la encuesta como más importantes fueron las que evaluaron con mayor severidad, mientras que las que consideraron menos importantes fueron evaluadas con mayor indulgencia (p. 480).

En cuanto a los estudios cuantitativos, la mayoría de los estudios han empleado MFRM para dar cuenta de los distintos efectos a que son sujetos los evaluadores. Dentro de estos efectos, el que mayor atención ha captado es el índice de severidad e indulgencia de los evaluadores. En el estudio que Schaefer (2008) llevó a cabo al analizar sesgos entre los correctores, el autor identificó que existía una relación entre el sesgo y subgrupos de evaluadores que desplegaban diferentes grados de severidad e indulgencia al evaluar textos. Puesto de otra manera, el autor identificó que cuando los evaluadores eran altamente severos en el contenido y/o la organización, contrariamente eran altamente indulgentes en el uso de la lengua y/o aspectos de la mecánica. Otro grupo, sorpresivamente, demostró patrones exactamente opuestos. Sin embargo, de acuerdo con Eckes (2012), uno de los problemas de Shaefer (2008) fue la incapacidad de explicar este extraño comportamiento de los evaluadores.

En un estudio posterior, Eckes (2012) abordó técnicas mixtas para dar cuenta de los patrones de sesgo en los evaluadores generados por sus estilos cognitivos. Según el

investigador, estos sesgos se dan por la manera mediante la cual los evaluadores interpretan y usan las categorías de la rúbrica, de acuerdo con su percepción sobre la importancia de cada una de éstas. Empleando datos de un estudio previo (Eckes, 2008b), en los cuales clasificó a los evaluadores de acuerdo con su estilo cognitivo, el autor empleó MFRM con el fin de identificar los sesgos presentados en grupos de evaluadores. Si bien es cierto que el autor identificó que los evaluadores se encontraban dentro del rango aceptable del uso de la rúbrica, éstos no eran particularmente dependientes unos de otros. El autor identificó que los evaluadores con estilos cognitivos similares (de acuerdo con el cuestionario que previamente se les había asignado) presentaban las mismas tendencias de evaluar más severamente las categorías que consideraban más importantes, mientras que la evaluación era más indulgente en aquellas concebidos de menor relevancia.

Según Myford (2012, p. 48), para dar respuesta al argumento de interpretación de Kane (2006), es necesario que se evidencien los siguientes aspectos relacionados con la cognición del evaluador:

- Los evaluadores emplean la rúbrica para otorgar un puntaje en cada criterio.
- Los evaluadores no permiten que criterios irrelevantes al constructo infieran en sus evaluaciones.
- Los evaluadores emplean las categorías de las rúbricas de manera consistente y precisa en cada evaluación.
- Los correctores evalúan cada texto de forma individual sin permitir que evaluaciones anteriores influyan en sus decisiones.
- Los correctores emplean de forma consistente el uso de la rúbrica sin alteraciones durante la sesión de evaluación.
- Los evaluadores son intercambiables; es decir, si un texto es evaluado por uno u otro corrector, el resultado será el mismo.

Finalmente, Myford (2012) sugiere que se sigan métodos cuantitativos y cualitativos que permitan comprender el fenómeno relacionado con el evaluador desde distintas aristas. Asimismo, menciona que es importante tener cuidado cuando se trata de estudios cualitativos con pocos evaluadores, puesto que éstos no pueden ser generalizables.

5.2.2.5.3. VARIABLES SUPERFLUAS QUE INFLUYEN EN LA APRECIACIÓN DEL TEXTO

Una de las problemáticas que se ha evidenciado en la corrección de textos se debe a que factores superficiales pueden influenciar negativamente la manera de evaluar (Rezaei y Lovorn, 2010). En el estudio que los autores condujeron, les asignaron a los evaluadores de un grupo un texto tipo ensayo para corregir, pero primero lo hicieron sin rúbrica y posteriormente, con ésta. Una de las características de los textos es que éstos se encontraban bien escritos en términos mecánicos: gramática, léxico, ortografía, pero no hacían referencia completa a los puntos planteados en la consigna. A un segundo grupo le asignaron también textos para evaluar de la misma manera que en el primero, pero en este caso, los autores insertaron deliberadamente 20 errores estructurales, de mecánica, ortografía y gramaticales. Sin embargo, los textos hacían clara referencia a los distintos puntos de la consigna. El objetivo de estos investigadores fue evidenciar si estos componentes de orden inferior de escritura influían negativamente en los resultados que los evaluadores otorgaban. Los investigadores encontraron que el uso de la rúbrica incrementaba considerablemente la varianza, cuando en realidad no debería haber sido así, pues en el caso del segundo ensayo, el único problema con éstos residía en la mecánica, por lo cual se esperaba que los resultados globales de las evaluaciones fueran más bien consistentes y elevados. Sin embargo, los investigadores notaron que muchos ensayos del segundo grupo (buen desarrollo de ideas, pero con errores mecánicos) habían recibido puntajes muy bajos incluso en el criterio del contenido y desarrollo de ideas, por lo cual era evidente que los errores en aspectos formales de la escritura habían influenciado la visión global de los evaluadores.

En el caso opuesto, los autores esperaban que en el primer grupo se asignaran valores altos a la cuestión mecánica y bajos al argumento dado por los estudiantes, pues éstos no atendían oportunamente cada uno de los aspectos de la consigna. Sin embargo, nuevamente, esto no fue así. De hecho, los correctores asignaban valores altos a categorías que ni siquiera figuraban en el texto como citar adecuadamente del texto fuente. Cabe resaltar que los investigadores intencionalmente asignaron a los aspectos mecánicos (gramática, ortografía y puntuación) solo un 10% del valor total de la rúbrica. No obstante, esta percepción sobre la “mala redacción del texto” influyó negativamente en la evaluación de otros criterios de la rúbrica bien elaborados. Lo anterior conlleva implicaciones en el uso de la rúbrica, pues es cierto que al ser usada para evaluar el desempeño escrito de estudiantes no hispanohablantes, es imprescindible considerar esta problemática al momento de evaluar. Hamp-Lyons (1991b) agrega que quienes escriben en segundas lenguas no necesariamente carecen de ideas ni de la mecánica para

transmitirlas, sino más bien de aquellos recursos lingüísticos necesarios para expresarse de manera eficaz. Por esto, Huang y Foote (2010) comentan que al contar con un mismo instrumento para evaluar el desempeño de los nativos y no nativos se pone en tela de juicio la equidad y justicia de la evaluación, puesto que los estudiantes no nativos siempre se verán perjudicados en su desempeño por aspectos que no son equiparables para ambos grupos lingüísticos. Los autores comentan que por este motivo se deberían construir rúbricas que hagan especial énfasis en el contenido y en la claridad para presentar los conceptos, seguidos –en segundo término– por aspectos lingüísticos y estructurales.

Wolfe y McVay (2012), por su parte, aseveran que existe otro efecto adicional, el cual denominan: diferencial de dimensionalidad. Según estos autores, éste ocurre cuando el evaluador se ve influenciado por variables externas al texto (p.ej., la caligrafía, el conocimiento del candidato o la percepción de mayor importancia sobre ciertos criterios de la rúbrica). Lai et al. (2015) mencionan que los evaluadores pueden aplicar erróneamente una rúbrica analítica por dos razones: por factores irrelevantes al ensayo (como la caligrafía) o porque los evaluadores no son capaces de distinguir las características del ensayo en las diversas categorías de la rúbrica. Este último factor es el que produce el efecto halo o de tendencia central entre los evaluadores. Los autores mencionan que una manera de mitigar este efecto se logra cuando un grupo de evaluadores evalúa solamente una categoría de la rúbrica en vez de todas. Del estudio que los investigadores condujeron, en el cual compararon la evaluación separada de cada una de las categorías de una rúbrica versus la evaluación de todas éstas, concluyeron que el efecto halo persiste cuando los correctores evalúan cada una de las categorías de la escala. El problema, sin embargo, radica en que se requiere de un vasto número de evaluadores y el proceso de jueceo resulta más costoso. En suma, todas estas variables ajenas al texto producen sesgos importantes que pueden conllevar a la falta de confiabilidad en los resultados que se le otorgan a un candidato o a un grupo de examinados.

5.3. MÉTODOS E INSTRUMENTOS

En este estudio segundo estudio, las inferencias que se establecieron fueron con respecto a la evaluación y a la generalización:

- **La rúbrica mide eficazmente las habilidades de escritura académica que requieren los estudiantes de posgrado no hispanohablantes.**

- **Los puntajes asignados por los evaluadores son confiables.**

Las aseercciones que se plantearon para respaldar estas inferencias fueron las siguientes:

Inferencia de evaluación

- La rúbrica alcanza los estándares de calidad determinados para MFRM.
- La percepción de los evaluadores sobre la rúbrica es positiva.
- La rúbrica mide eficazmente las habilidades de escritura académica requeridas a nivel posgrado.

Inferencia de generalización

- Los puntajes asignados por los evaluadores alcanzan los estándares de calidad determinados para MFRM.

Para poder dar evidencia que permitiera asentir las aseercciones previamente enunciadas, este segundo estudio se conformó de un estudio mixto dividido en tres fases: la calidad del proceso de evaluación mediante MFRM, la percepción de los evaluadores respecto de la rúbrica y la factibilidad de la rúbrica de medir las habilidades de escritura académica requeridas a nivel posgrado. Debido a que tanto las tareas de la prueba escrita como las rúbricas para evaluarlas eran de diseño reciente, la primera fase se dividió en dos etapas: pilotaje con la primera versión de la rúbrica y corrección con la segunda versión de la rúbrica.

5.3.1 LA CALIDAD DEL PROCESO DE EVALUACIÓN

De acuerdo con el argumento de validación de la PE, en este apartado se formularon dos inferencias sobre la calidad en el proceso de corrección de la PE del EXELEAA. La primera inferencia se relaciona con la calidad de la rúbrica analítica, y la segunda, con la calidad de los evaluadores. Debido a que el proceso de asignación del puntaje a un texto se encuentra íntimamente relacionado entre la representación mental del evaluador sobre el texto y sobre la rúbrica, ambas inferencias se abordan en este apartado. Esta fase se llevó a cabo en dos partes. Debido a que la rúbrica analítica fue diseñada para este nuevo examen, primero se llevó a cabo un pilotaje con la rúbrica preliminar. Una vez afinada la rúbrica con base en la experiencia de los evaluadores y de los análisis estadísticos preliminares, se llevó a cabo la corrección de los textos con la rúbrica modificada.

5.3.1.1. PRIMERA PARTE: PILOTAJE CON LA PRIMERA VERSIÓN DE LA RÚBRICA

5.3.1.1.1. LA MUESTRA

En un principio se planeó contar con una muestra de textos pertenecientes exclusivamente a estudiantes del posgrado. Sin embargo, debido a que esta población no hispanohablante es escasa y la gran mayoría de los estudiantes ya cuentan con un certificado de español, pues ya están inscritos en el posgrado, la respuesta en las primeras convocatorias fue baja (2014-1 y 2015-1). Por esta razón, se decidió ampliar la muestra e incorporar en los textos a evaluar también a alumnos de licenciatura no hispanohablantes. Para el pilotaje, se contó con una muestra de 37 estudiantes, 19 de posgrado y 18 de licenciatura.

5.3.1.1.2. LAS TAREAS DE LA PE

En la PE del examen se les pidió a los sustentantes que escribieran dos textos: descripción y contraste de gráficos, y un ensayo argumentativo (ver Anexo 2.1). El modo de administración fue en papel, debido a que la plataforma aún no se encontraba habilitada para que los textos fueran almacenados en el sistema.

5.3.1.1.3. LA PRIMERA VERSIÓN DE LA RÚBRICA ANALÍTICA

La rúbrica analítica fue construida específicamente para la evaluación de las tareas concebidas para el EXELEAA (ver Anexo 2.2). En la introducción se enunciaron aquellos componentes que fueron incluidos en la rúbrica analítica; sin embargo, ahora se hará una descripción detallada de cada uno de éstos. Siguiendo el modelo de competencia comunicativa de Canale y Swain (1980), se incluyeron cuatro componentes en la rúbrica inicial: sociolingüístico, pragmático-discursivo, lingüístico y estratégico.

En la primera sección, componente lingüístico, se evaluó la habilidad de los examinados para emplear las convenciones académicas formales propias de ensayos argumentativos y de la descripción de información visual. En esa sección también se evaluó si el discurso cumplía o no con su propósito comunicativo.

En la segunda sección, componente pragmático-discursivo, se incluyeron tres categorías: el desarrollo de ideas, coherencia y cohesión. El desarrollo de ideas para la tarea 1 evaluaba la descripción y contraste de información visual, así como la habilidad de sintetizar y resumir información. El desarrollo de ideas para la tarea 2 incluía la habilidad de escribir un ensayo

argumentativo bien desarrollado con los siguientes componentes: introducción, enunciación de la tesis, desarrollo de argumentos con ideas de apoyo basadas en razones o ejemplos, el reconocimiento de contraargumentos y la habilidad para refutarlos. Finalmente, también se consideró la inclusión de una conclusión que sintetizara la problemática y reafirmara la tesis.

Para evaluar la cohesión, se consideró la microestructura a nivel oracional mediante la relación de palabras dentro de una oración (i.e., concordancia entre sustantivos y adjetivos, deícticos, elipsis, anáforas y catáforas) y también entre oraciones dentro de un mismo párrafo. La coherencia englobaba también las relaciones semánticas generales del texto a través del uso de marcadores discursivos entre oraciones y entre párrafos.

La tercera sección, competencia lingüística, incluía cinco categorías: léxico, gramática, sintaxis, puntuación y ortografía. La categoría del léxico evaluaba la selección apropiada de palabras y de vocabulario dependiendo del tipo de tarea y del tema asignado. La precisión gramatical incluía el uso correcto de categorías morfológicas, tiempos verbales y modos (indicativo vs subjuntivo). La complejidad sintáctica comprendía un amplio rango de oraciones subordinadas y marcadores subordinantes. Finalmente, la puntuación y la ortografía cubrían la habilidad del examinado de usar un amplio rango de signos de puntuación y de la precisión en la ortografía y el correcto uso de acentos.

Finalmente, la cuarta sección, competencia estratégica, evaluaba la evidencia del uso de diferentes estrategias (p.ej., planeación del escrito mediante un esquema o el monitoreo del texto a través de autocorrecciones).

5.3.1.1.4. LOS EVALUADORES

En esta fase de pilotaje participaron cinco evaluadores. Cuatro de los evaluadores contaban con una amplia experiencia en la enseñanza de español como lengua extranjera y también en el diseño y uso de rúbricas para evaluar la expresión escrita. Una evaluadora era novata, con formación en pedagogía y sin experiencia docente ni en evaluación mediante rúbricas.

5.3.1.1.5. EL PROCESO DE CORRECCIÓN

En esta etapa de pilotaje, los cinco evaluadores corrigieron los dos textos producidos por 37 estudiantes (ver la matriz de evaluación en el Anexo 2.3). Cada alumno, en sus dos tareas,

fue evaluado por dos correctores. Como se puede apreciar, la mayoría de los textos fueron corregidos por el evaluador 1 y 2. Al inicio de la sesión de evaluación, se proporcionó una breve inducción sobre el manejo de la rúbrica. La evaluación se llevó a cabo de manera colectiva con el fin de poder discutir posibles dudas o dificultades que emergieran en el uso de la rúbrica o para la asignación de un puntaje.

5.3.1.1.6. EL ANÁLISIS ESTADÍSTICO EN MFRM

Para poder evidenciar la calidad de la rúbrica y de los evaluadores, se empleó el análisis estadístico MFRM, mediante el programa estadístico FACETS versión 3.71.4 (2015), desarrollado por John Linacre (1989). Este programa es una extensión del modelo de Rasch y permite segmentar diversas facetas que puedan estar involucradas como posibles fuentes de variabilidad. En el caso de las PE de respuesta construida, estas variables se encuentran en el sustentante, la tarea, el evaluador y los criterios de la rúbrica. En el caso de la tarea, el programa permite evidenciar si una tarea resulta más difícil que otra. Con relación al evaluador, se pueden establecer los índices de severidad e indulgencia de éstos. Finalmente, en cuanto a la rúbrica, se puede determinar el funcionamiento y la severidad con que se aplican cada una de las categorías establecidas en ésta. Finalmente, el programa permite calcular un puntaje justo para el examinando, eliminando los posibles efectos de variabilidad producidos por las tres facetas antes mencionadas.

5.3.1.1.6.1. EL PROGRAMA ESTADÍSTICO FACETS

Dado que el análisis de MFRM se empleó para evaluar la calidad del proceso de corrección de la prueba escrita (i.e., la calidad de la rúbrica y de los puntajes de los evaluadores), se hará una detalla descripción de su funcionamiento y de lo que se puede obtener de su análisis.

El programa que se empleó se llama FACETS⁸. Este programa permite utilizar el modelo de Rasch para aquellos casos en los que se cuenta con facetas múltiples. En el caso de las PE, las facetas que se presentan son: el nivel del examinado, el evaluador, la tarea y las categorías de la rúbrica analítica. El análisis es una extensión del modelo de crédito parcial de Rasch y su característica principal es que permite la inclusión de ítems politómicos, mientras que el Rasch solamente aquellos que son dicotómicos (Linacre, 2012). Entonces, el modelo permite examinar

⁸ Consúltense Linacre 2012a, 2012b, 2012c, 2012d, 2013) para información detallada del funcionamiento de FACETS.

hasta qué punto las variables (consideradas como facetas) contribuyen al error de medida. Consecuentemente, la característica más sobresaliente de este modelo consiste en que los puntajes de los sustentantes no dependen del evaluador ni de la tarea: objetividad específica (Bond y Fox, 2007; Eckes, 2009a; Engelhard, 2008b; Prieto, 2011). Eckes (2009a, p. 4) menciona que esta objetividad específica se refiere a que las mediciones del examinado se encuentran libres del examen; es decir, que cuando las observaciones en un modelo de Rasch son lo suficientemente precisas, la medición de los examinados es invariable a las tareas o evaluadores. Para que el programa funcione adecuadamente, no se requiere, entonces, que los evaluadores presenten puntajes exactos, sino más bien consistentes. El programa calibra las distintas facetas de manera simultánea y otorga un resultado equivalente en una escala de intervalo medida en lógitos (*log-odd units*)⁹. La ventaja de esta escala es que se pueden comparar las distintas facetas en términos de su dificultad o severidad.

El análisis de facetas múltiples con el programa MFRM ha sido empleado en numerosas áreas del saber (Bond y Fox, 2007), sobre todo en las ciencias médicas (Linacre, 2013) y en la educación, en la evaluación de textos, en donde es indispensable que las evaluaciones sean consistentes y confiables. De hecho, en el ámbito educativo en general, este método se ha empleado extensivamente durante la última década para ver cómo se comportan las diversas variables que inciden en el puntaje que se le otorga a un sustentante. Estas variables generalmente se relacionan con el evaluador, la tarea y las categorías de la rúbrica – en caso de que estas sean analíticas –.

Autores como Eckes (2009a), Knoch (2007, 2009), Prieto (2011), Prieto y Nieto (2014) y Wind y Engelhard (2013) han empleado el análisis de MFRM para evidenciar el funcionamiento de las variables de las PE de los exámenes de lenguas administrados a gran escala. Este tipo de estudios es de especial relevancia puesto que permiten determinar, por una parte, cuál es el funcionamiento de una rúbrica, mientras que por otra, qué evaluadores son más severos o indulgentes al momento de evaluar. Asimismo, es una herramienta útil pues permite otorgarle al

⁹ El lógito es el logaritmo natural que se obtiene de la división entre dos probabilidades. En el caso de ítems politómicos (como es el caso de la rúbrica escrita aquí empleada con valores de 0, 1, 2, 3), este logaritmo representa la probabilidad de obtener un puntaje y el puntaje inmediatamente inferior. A diferencia de la Teoría Clásica, en la cual se emplean los resultados netos para estimar la habilidad del sustentante, en los modelos de Rasch se emplea una transformación logística que convierte los resultados netos en una escala de log-odd units. Esta escala es dependiente de otras como la severidad del evaluador, la dificultad de la tarea o de los criterios de la escala (Engelhard y Wind, 2013).

sustentante un puntaje justo que depende de sus habilidades y no de quien lo evalúa. Evidentemente, gracias a estos resultados, se pueden tomar medidas para darle retroalimentación a los evaluadores o en determinados casos, cuando su evaluación no es consistente, sustituirlos por otros. De hecho, las variables de la rúbrica y de los evaluadores van de la mano, y por ello, es indispensable garantizar primeramente que la rúbrica mida lo que deseamos medir. Como se pudo ver en el apartado anterior, si la rúbrica no es clara ni precisa, los evaluadores presentarán dificultades al emplearla, lo cual se acentuará más con las características propias que determinarán su grado su severidad o indulgencia al evaluar los textos. Sin embargo, el hecho de que la evaluación sea propensa a errores, sobre todo en el caso de la inconsistencia de los evaluadores, pone en tela de juicio la validez de los exámenes.

El programa FACETS presenta múltiples ventajas de aplicación. En primera instancia, este permite calibrar los valores de los sustentantes, los evaluadores, las tareas y las categorías de la rúbrica. Además de esta ventaja, Eckes (2009a), Knoch (2007), Myford y Wolfe (2003, 2004) y Prieto (2011) mencionan que otro de los grandes potenciales del MFRM es que éste nos permite identificar diversos efectos del evaluador: grado de severidad, efecto halo y el efecto de la tendencia central. Asimismo, otra de las ventajas del modelo es que permite determinar el grado en que las calificaciones otorgadas por un evaluador correlacionan o son consistentes con las de los demás evaluadores (correlaciones inter-evaluador). Además, FACETS permite identificar la confiabilidad de las calificaciones de las distintas facetas al calcular cuál es la proporción de la varianza verdadera respecto de la varianza observada de las medidas (separación de las medidas). En el caso de los evaluados, se espera que este valor sea alto, pues los sustentantes presentan niveles desiguales, mientras que, en el caso de los correctores, se espera que este valor sea reducido, debido a que la calificación otorgada por dos o más evaluadores a un mismo sustentante debería ser lo más cercana posible. Finalmente, una de las ventajas más importantes de FACETS es que permite trabajar con datos incompletos (Eckes, 2009a; Esfandiari y Myford, 2013; Linacre, 2012b, 2013). En este caso, resulta relevante, pues no todos los evaluadores corrigieron los mismos ensayos, por lo cual otros modelos estadísticos como la Teoría de la Generalizabilidad (conocida como Teoría G) o la Teoría Clásica resultarían ineficientes.

Por ejemplo, en el caso del estudio de Esfandiari y Myford (2013) un grupo de seis docentes evaluaron 188 ensayos, mientras que los estudiantes, en una evaluación de pares, solamente evaluaron dos ensayos cada uno; sin embargo, dado que el modelo de MFRM funciona

adecuadamente con datos faltantes, el anclaje de los estudiantes que evaluaron los ensayos de dos de sus compañeros se pudo efectuar con los seis docentes quienes evaluaron todos los textos.

Rating data that originate from incomplete block designs can prove to be very challenging for researchers to analyze, particularly if they try to use common classical theory-based approaches, such as analysis of variance or generalizability theory, given their underlying assumptions. However, the Many-facet Rasch measurement approach is well suited to analyzing rating data from incomplete blocks designs if the judging plan that was implemented results in sufficient linkages. To be sufficiently linked, each element of each facet... must be able to be linked either directly or indirectly to every other element (Esfandiari y Myford, 2013, p. 121).

5.3.1.1.6.2. ANTECEDENTES DE MFRM

El modelo de MFRM (Linacre, 1989) conceptualiza los resultados observados como la interacción entre los distintos elementos de las facetas (p.ej., el examinado 1, el evaluador A, la tarea C, la categoría de la rúbrica 7). Asimismo, el modelo presupone a los evaluadores como expertos independientes y no como máquinas, por lo cual un cierto grado de severidad e indulgencia entre los evaluadores es comprensible. Este modelo encuentra sus fundamentos en el modelo de Rasch. George Rasch (1960, en Linacre 2013) generó un modelo matemático para las ciencias sociales en el cual podemos establecer una concatenación de niveles de dificultad en un examen. Este modelo se basa en ítems dicotómicos basados en la probabilidad de acertar o fallar en un ítem. Cuando la probabilidad de un examinado de acertar un ítem es 0.5, entonces se dice que se ha ubicado su nivel de habilidad. Es decir, su probabilidad de acertar los ítems relacionados con su índice de dificultad es de un 50 por ciento. Entonces, su probabilidad de acertar ítems más fáciles será mayor y su probabilidad de acertar ítems más difíciles, será menor.

El modelo de Rasch fue modificado en 1978 por David Andrich, conocido como Rasch-Andrich Rating Scale Model (Linacre, 2013). Este investigador propuso que en el uso de una escala, los puntajes se encuentran concatenados de tal suerte que éstos serían semejantes a la suma de categorías dicotómicas. Entonces, este autor diseñó un modelo que especifica la probabilidad de que una persona con una cierta habilidad B_n sea observada con un puntaje determinado “j” en un ítem politómico (p.ej., escala tipo Likert) con cierta dificultad D_i y la probabilidad de que sea observada en el puntaje inmediatamente inferior $j-1$.

Este modelo fue nuevamente mejorado por Masters (1982) quien propuso que los examinados deberían recibir créditos parciales en sus respuestas, dado que había distractores

que eran parcialmente correctos (Linacre, 2013). Así que, este investigador dio pie al *Rasch-Masters Partial Credit Model*. Esta forma de concebir los puntajes en un ítem particular fue de utilidad en rúbricas en donde cada una de las categorías, aunque relacionadas entre sí, son independientes unas de otras. Entonces, esta noción de crédito parcial se extrapoló a los posibles puntajes que un evaluador puede otorgar a una categoría específica (p.ej., la gramática en una rúbrica analítica).

5.3.1.1.6.3. DESCRIPCIÓN DEL MODELO MFRM

El modelo de Linacre (1989), a su vez, representa una extensión del modelo de crédito parcial de Rasch-Masters¹⁰. No obstante, su peculiaridad reside en que se pueden contemplar distintas facetas: el examinado, el juez o evaluador, la tarea y las categorías de la rúbrica analítica (en caso de PE). En este modelo, el lógito corresponde a la probabilidad de que una persona reciba una calificación en una tarea, en una categoría, por un evaluador, entre la probabilidad de que reciba una calificación inmediatamente inferior. $P_{nmjk} / P_{nmj(k-1)}$ (Linacre, 2013). La ecuación que se desprende es la siguiente:

$$\log (P_{nmjk} / P_{nmj(k-1)}) = B_n - A_m - D_i - C_j - F_k$$

Donde:

P_{nmjk} = la probabilidad de que una persona en la tarea m , sea evaluada en el ítem i por el evaluador j y reciba la calificación k .

$P_{nmj(k-1)}$ = la probabilidad de que una persona en la tarea m , sea evaluada en el ítem i por el evaluador j y reciba la calificación inmediatamente inferior ($k-1$).

B_n = la habilidad de la persona (p.ej., alumno 1)

A_m = el reto de la tarea (p.ej., ensayo)

D_i = la dificultad de la categoría de la rúbrica (p.ej., ortografía)

C_j = el nivel de severidad del evaluador (p.ej., evaluador 1)

F_k = el límite entre recibir el puntaje k y el puntaje $k-1$ (p.ej., 2 y 1)

De manera más puntual, el lógito se refiere a la probabilidad ($P_{nmjk} / P_{nmj(k-1)}$) de que una persona con una cierta habilidad, B_n , al ser observada tras desempeñar una tarea con un

¹⁰ FACETS Resulta útil en la evaluación del desempeño llevada a cabo por un par o grupo de evaluadores. Linacre (2013) menciona que es ideal para evaluar ensayos, portafolios y otro tipo de evaluación de desempeño en la cual intervienen jueces.

reto o dificultad específica, A_m , sea evaluada en un ítem con una cierta dificultad, D_i , por un evaluador con determinado grado de severidad, C_j , reciba una determinada puntuación " k " (Linacre 2013, p. 13). Para ejemplificar cómo se leería esta ecuación en una PE, se podría plantear de la siguiente manera: La probabilidad de que el alumno Jorge (n , persona) sea observado en el escrito de un ensayo (m , tarea), y sea evaluado en ortografía (i , categoría de la rúbrica) por el evaluador Andrés (j , juez o evaluador) con un puntaje de 2 (k , escala de la rúbrica en rangos de 0 a 4 puntos).

En esta ecuación, el lógito es la variable dependiente, mientras que las diversas facetas (nivel de la persona, dificultad del ítem y severidad del evaluador) corresponden a la variable independiente. En este modelo se asume que los pasos (F_k) pueden variar entre los evaluadores. La escala del lógito puede oscilar entre 0 y $\pm \infty$. El cero representa el punto medio de cada una de las facetas, aunque cada una de éstas puede variar libremente de forma independiente dependiendo de la dificultad de la tarea, la severidad del evaluador y de los criterios de la rúbrica. El lógito representa un error típico de medida además de los índices de ajuste entre las respuestas observadas y las predichas por el modelo. Los estadísticos obtenidos pueden ser a nivel faceta o también a nivel grupal (Myford y Wolfe, 2004).

5.3.1.1.6.4. CRÍTICAS AL MODELO MFRM

Wang y Yao (2012) mencionan que es mejor emplear otros modelos multifacetas en vez de MFRM, dado que, según estos autores, uno de los problemas del programa FACETS para llevar a cabo el análisis estadístico de MFRM se debe a la presencia de sesgos asintomáticos causados por el método de verosimilitud máxima (incondicional) conjunta (p. 1). Los autores mencionan que el modelo de Yao¹¹ es una extensión y mejoramiento del de Linacre (1989), puesto que puede funcionar con ítems de respuesta construida y también de opción múltiple simultáneamente. Asimismo, este modelo permite incorporar múltiples parámetros y un atributo latente o múltiples atributos latentes. Entonces, la ventaja de este modelo sobre el MFRM subyace en la posibilidad de comparar ítems de respuesta construida con ítems de opción múltiple. Según los autores, cuando los paquetes de análisis de multifacetas solamente incluyen ítems de respuesta construida, se reduce la confiabilidad del examen y se incrementan los sesgos en el resultado de las habilidades del examinado.

¹¹ Consúltese Yao (2010). *Bayesian Multivariate Item Response Theory (Computer program)*. Disponible en el portal: <http://www.bmirt.com/6271.html>
Última consulta, 30 de diciembre de 2016.

El estudio que aquí se presenta contiene exclusivamente una sección de escritura con dos pruebas de respuesta construida, razón por la cual se empleó el método de Linacre: MFRM, mediante el paquete estadístico FACETS. Sin embargo, en los casos en los cuales se desea contrastar pruebas de respuesta construida con aquellas de opción múltiple, sería prudente el uso del modelo de Yao.

5.3.1.2. SEGUNDA PARTE: CORRECCIÓN CON LA SEGUNDA VERSIÓN DE LA RÚBRICA

5.3.1.2.1. LA MUESTRA

Dado que el programa estadístico que se empleó para evaluar los textos funciona con muestras relativamente grandes –por lo menos 100 candidatos–, se decidió incrementar dicha muestra a toda la población universitaria no hispanohablante de la UNAM y de otras universidades privadas del área metropolitana y también de la Universidad de Guadalajara. La muestra se recolectó del período académico 2014-1 a 2015-2. El motivo por el cual desde un inicio se pensó en descartar a los estudiantes de licenciatura se debió fundamentalmente a que esta población generalmente se encuentra en la UNAM por un semestre de intercambio universitario y muchas veces no cuentan con experiencia de escritura académica en español hasta que termina su semestre.

Entonces, la muestra final de 100 participantes se conformó por: 45 alumnos no hispanohablantes del posgrado de la UNAM (conformados en dos sesiones diferentes), 18 de licenciatura de intercambio en la UNAM, 25 de licenciatura de intercambio en la Universidad de Guadalajara, 5 de licenciatura del Instituto Tecnológico de Estudios Superiores de Monterrey, Campus Ciudad de México y 7 de licenciatura de la Universidad Panamericana. El Cuadro 5.3 presenta un extracto de la muestra que en conjunto participó en la PE del EXELEAA. En el Anexo 2.4 se ofrece un desglose de cada uno de los participantes.

5.3.1.2.2. LAS TAREAS DE LA PE

En la PE del examen se les pidió a los sustentantes que escribieran dos textos: descripción y contraste de gráficos, y un ensayo argumentativo (ver Anexo 2.1). Esta primera versión fue administrada a una población de 68 estudiantes. Sin embargo, con base en la experiencia al evaluar los textos con la primera rúbrica, así como del análisis del apartado teórico, se diseñó una nueva propuesta que fue aplicada a los 32 estudiantes restantes (ver Anexo 2.5).

Cuadro 5.3: Datos de los examinados

Universidad Nacional Autónoma de México	Universidad de Guadalajara	Instituto Tecnológico de Estudios Superiores Monterrey	Universidad Panamericana
63	25	5	7
45 posgrado	55 licenciatura		
Nacionalidad	francesa	16	
	alemana	15	
	estadounidense	14	
	italiana	8	
	japonesa	8	
	checa	3	
	canadiense	3	
	belga	3	
	austriaca	3	
	brasileña	3	
	china	3	
	inglesa	2	
	taiwanesa	2	
	otras nacionalidades	7	
sin datos	10		
Total			100

Asimismo, cabe aclarar que el modo de administración difirió en algunos casos. 52 estudiantes escribieron sus textos en computadora, mientras que los otros 48 lo hicieron en papel. Esto se debió a que en un inicio la plataforma del examen no se encontraba aún habilitada para que los textos fueran almacenados en el sistema. De igual manera, en el caso de los exámenes administrados en las escuelas privadas del área metropolitana participantes, éstos también se escribieron a mano. Sin embargo, cabe aclarar que en todos los casos se administró la prueba bajo las mismas condiciones de tiempo y también sin el apoyo de diccionarios, traductores o correctores automáticos de texto.

5.3.1.2.3. LA SEGUNDA VERSIÓN DE LA RÚBRICA ANALÍTICA

Al final de las sesiones de evaluación de la etapa del pilotaje, se llevaron a cabo talleres con los cinco evaluadores con el fin de homologar criterios, identificar textos prototípicos de niveles de desempeño y dar retroalimentación sobre el uso de la rúbrica. Posteriormente, la responsable del examen y el investigador revisaron nuevamente la rúbrica y, tomando en cuenta los estadísticos preliminares así como las observaciones de los evaluadores y la experiencia durante el pilotaje, enmendaron la rúbrica y elaboraron una segunda versión (ver Anexo 2.6).

5.3.1.2.4. LOS EVALUADORES

En esta etapa, cada uno de los 100 estudiantes fue evaluado de forma independiente por 6 correctores humanos. Cabe mencionar que la mayoría de los candidatos respondió ambas tareas, aunque hubo quienes por razones desconocidas únicamente respondieron a una de ellas (26 estudiantes). A estos últimos se les evaluó su desempeño en el presente estudio exclusivamente con una tarea. En esta fase de evaluación participaron 6 evaluadores: 2 experimentados, la responsable del examen y el investigador, y 4 novatos a quienes se les remuneró económicamente para poder evaluar los textos. Lo anterior debido a que quienes participaron en el pilotaje no les fue posible seguir colaborando en el proyecto. Los evaluadores novatos no contaban con experiencia alguna en la evaluación de escritos de exámenes de español como lengua extranjera mediante rúbricas. Solamente la evaluadora 4 estaba familiarizada con la evaluación de PE, pero para evaluar el desempeño de nativohablantes.

5.3.1.2.5. EL PROCESO

En una primera instancia, los dos evaluadores experimentados corrigieron los textos de los 100 estudiantes en ambas tareas. Esto con el fin de poder aclarar dudas y discrepancias surgidas durante la evaluación. Debido a restricciones económicas, los otros cuatro evaluadores también evaluaron a los 100 estudiantes, aunque solamente en una de las dos tareas. Entonces, cada estudiante fue evaluado por cada uno de los 6 correctores en una o ambas tareas. Sin embargo, para lograr la mayor cantidad de interacciones posibles, se buscó la alternancia entre evaluadores y tareas (ver matriz de evaluación en el Anexo 2.7).

El proceso de capacitación se llevó a cabo de forma independiente con cada uno de los cuatro evaluadores novatos de la siguiente forma. A los evaluadores se les dio la nueva rúbrica junto con los dos textos elaborados por cuatro examinados, mismos que habían sido previamente seleccionados durante el pilotaje como ejemplos prototípicos de los distintos niveles de desempeño. Se les pidió que estudiaran la rúbrica y, en caso de haber dudas, éstas se explicaron y aclararon. Una vez que se sentían familiarizados con la rúbrica, se les pidió que evaluaran los 8 textos (gráfica y ensayo argumentativo de cada sustentante) también de forma independiente. Posteriormente, se les pidieron sus resultados para poder cotejarlos con los puntajes previamente establecidos. En este momento, de manera individual, se les proporcionó retroalimentación y se clarificaron las dudas con respecto a la asignación de las distintas categorías de la rúbrica. Estas sesiones de retroalimentación se llevaron a cabo de forma presencial en un lapso de aproximadamente una o dos horas para cada evaluador. Concluida la

sesión de retroalimentación, se les otorgaron los textos distribuidos en tres paquetes para que los evaluaran a lo largo de un mes. Cabe aclarar que los paquetes dos y tres se entregaron una vez que habían concluido con el anterior. Esto para poder comparar los resultados y proporcionar retroalimentación adicional en caso de que fuera necesario.

5.3.1.2.6. EL ANÁLISIS

El procedimiento para analizar la calidad del proceso de corrección fue el mismo que para la etapa del pilotaje. Se utilizó el paquete FACETS versión 3.71.4 (2015).

5.3.2. LA PERCEPCIÓN DE LOS EVALUADORES SOBRE LA RÚBRICA

Una vez que concluyó el proceso de corrección, se le preguntó al segundo grupo de evaluadores novatos su opinión sobre la rúbrica analítica modificada y también sobre el monitoreo y retroalimentación que recibieron durante el proceso de corrección.

5.3.2.1. LOS PARTICIPANTES

Debido a que los dos evaluadores experimentados participaron en el proceso de modificación de la rúbrica analítica, la participación se restringió a los cuatro evaluadores novatos.

5.3.2.2. EL INSTRUMENTO

Para poder conocer las impresiones de los evaluadores novatos con respecto a la rúbrica modificada, se diseñó un cuestionario abierto en *google forms* (ver Anexo 2.8) que se les envió a los evaluadores una vez que terminaron el proceso de corrección de los textos. Las preguntas fueron diseñadas para conocer tres aspectos: la formación académica de los evaluadores, su percepción de la rúbrica y su opinión sobre la retroalimentación recibida durante el proceso de corrección. La información se almacenó y organizó en una hoja de EXCEL.

5.3.3. EL ALCANCE DE LA RÚBRICA DE LA PE PARA EVALUAR HABILIDADES DE ESCRITURA ACADÉMICA

Debido a que las rúbricas se diseñan con base en el constructo, pero también enfocadas al propósito de la evaluación, es importante verificar si éstas miden las habilidades que se requieren en el “mundo real”. En esta etapa, se contrastaron las categorías de la rúbrica con las habilidades de escritura académica que los docentes del posgrado consideraron como

importantes para el éxito académico de los estudiantes de este ciclo universitario (consúltese el Capítulo 4, Estudio 1 de esta investigación).

5.3.3.1. LOS PARTICIPANTES

Se convocó un panel de cinco expertos en enseñanza de español como lengua extranjera y en la evaluación y diseño de rúbricas para evaluar la producción escrita en tareas diseñadas para el ámbito académico.

5.3.3.2. LOS INSTRUMENTOS

El cotejo se realizó con la lista de descriptores de las habilidades de escritura académica (ver Anexo 1.3) y la rúbrica analítica modificada para evaluar los textos de la PE (ver Anexo 2.6).

5.3.3.3. EL PROCESO

Para poder delimitar el alcance de la PE del EXELEAA, se empleó un procedimiento de validación de expertos (véase Becker, 2011 y Rosenfeld et al., 2004). Siguiendo la línea de estos autores, se les pidió a los expertos que indicaran, según su juicio, si cada uno de los 33 enunciados de tareas de escritura contemplados en la encuesta tipo Likert eran evaluados mediante las categorías de la segunda versión de la rúbrica analítica. Debido a que el interés era que los docentes no se vieran influenciados por las opiniones de los demás, este trabajo de cotejo se llevó a cabo de manera independiente. A cada uno de los especialistas se le entregó un documento que contenía los descriptores de escritura y la rúbrica. Junto a cada uno de los descriptores, los expertos debían identificar si éste era evaluado o no mediante la rúbrica, según su juicio. Una vez que concluyeron, estos documentos le fueron devueltos al investigador.

5.4. RESULTADOS

En esta sección se presentan los resultados de la calidad del proceso de evaluación. Como preámbulo, se dan a conocer algunas de las variables que se identificaron en el tipo de tarea y el modo de administración. Posteriormente, se describen los resultados estadísticos proporcionados por MFRM con respecto a las tres facetas de variabilidad: la tarea, la rúbrica y los evaluadores. Más adelante se dan a conocer los resultados del cuestionario sobre la percepción de los evaluadores. Finalmente, se muestran los resultados del cotejo del juicio de expertos en relación con las habilidades académicas que mide la rúbrica analítica de la PE del EXELEAA.

5.4.1. LAS TAREAS DE LA PRUEBA ESCRITA

Como se mencionó en el apartado anterior, la PE consistió de dos tareas: la descripción y contrastación de gráficos, y la redacción de un ensayo argumentativo (ver Anexo 2.1). Sin embargo, con base en la experiencia del pilotaje y de la revisión de la literatura, se hizo una segunda versión de las tareas (ver Anexo 2.5).

A continuación, se reportan los rasgos más sobresalientes de las dos tareas discretas incluidas en el pilotaje de la PE EXELEAA. En el caso de la gráfica, el primer problema que identificamos se debía a la interpretación de ésta, puesto que varios estudiantes escribieron que los datos no se podían comparar debido a que no se contaba con la información de los hijos, lo cual impedía que se pudiera contrastar eficazmente el logro educativo de las mujeres con sus padres.

Ejemplos de problemas evidenciados para contrastar las gráficas.

- “De manera general, se observa un incremento del nivel de educación entre estas dos generaciones. Pero no nos podemos dar cuenta de la verdadera diferencia entre los hombres y las mujeres, dado que no sabemos la proporción de hijos escolarizados, etc.”.

En segundo lugar, cabe mencionar que algunos estudiantes retomaron paquetes léxicos de la consigna, tanto de las gráficas como del ensayo; aun cuando este input era mínimo. Esto demuestra que cuando el estudiante posee fuentes escritas es frecuente que copie fragmentos del texto, lo cual resulta difícil de identificar para un evaluador humano.

Ejemplo de calcos literales de las tareas 1 y 2. Se destaca en negritas lo que el alumno copió literalmente de la información proporcionada en la consigna.

- “**Vimos dos gráficas sobre el logro educativo que han alcanzado en la última década las hijas con respecto a sus padres en un país de América Latina**”.
- “**En estos dos gráficos recibimos información sobre el logro educativo que han alcanzado las hijas con respecto a sus padres**”.
- “**Las gráficas muestran el logro educativo de las hijas con respecto a sus padres en un país de América Latina**”.
- [Milosavljevic dijo “**Si bien la educación por sí misma amplía los horizontes de los individuos y constituye un beneficio para éstos, la situación educativa de hombres y mujeres tiene diferentes beneficios para unos y otros y para sus hijos e hijas**”].

- [Según Milosavljevic (2007: 96) **“Si bien la educación por sí misma amplía los horizontes de los individuos y constituye un beneficio para éstos”**].

Otra de las dificultades observadas en el ensayo argumentativo propuesto se debió a que en realidad la consigna no permitía que los estudiantes utilizaran argumentos en contra de la educación de las mujeres, dado que esto representaría una postura desigual de género. Sin embargo, dentro de la consigna se especifica que mencionen contraargumentos y en la rúbrica se los evalúa.

Ejemplo de la problemática de evidenciar contraargumentos:

- “Milosavljevic dice aquí que la educación conlleva beneficios diferentes para hombres y mujeres pero yo no veo diferencias de género en los beneficios de la educación”.
- “Tal vez, se dice que la situación educativa tiene diferentes beneficios para los hombres y las mujeres, no obstante, yo considero que la educación es importante para todos”.
- “[...] estoy muy sorprendida cuando se dice que “la situación educativa de hombres y mujeres tiene diferentes beneficios para unos y otros y para sus hijos e hijas” todo el mundo debería tener los mismos resultados y beneficios”.

La otra problemática que se evidenció en la propuesta inicial de las tareas 1 y 2 es que éstas se encontraban ligadas; es decir, ambas tareas versaban sobre la educación de las mujeres. Además, en la consigna de la tarea 2 se explicita que el sustentante puede retomar información de la tarea 1. En realidad, ambas tareas deberían ser independientes una de la otra, pues además eso coadyuvaría a que los estudiantes explotaran al máximo su repertorio léxico en dos temáticas diferentes.

Finalmente, cabe destacar que la escritura a mano dificultó considerablemente la lectura de los textos, ya que en múltiples ocasiones el evaluador requería interpretar lo que el estudiante había escrito. Cuando esto sucede, es recurrente que el corrector se brinque los segmentos en los cuales no se puede comprender el texto, lo cual deja huecos en las ideas plasmadas en el escrito. En contraparte, es cierto que cuando el texto se escribe a mano se ve mucho más largo de lo que en realidad es, lo cual visualmente impacta en el evaluador como si el desarrollo hubiera sido más sustancial, aunque en realidad se están valorando aspectos superficiales relacionados con una mera apreciación visual. Además, es cierto que la escritura a mano se percibe como si fuera un primer borrador, mientras que los errores en computadora destacan más, pues uno está acostumbrado a ver textos en computadora como productos finales. Otro

impacto visual es que la letra que es más legible tiende a incidir positivamente en cómo percibe el evaluador el texto de entrada. Pareciera entonces que el contenido del texto es también, por ende, más claro. A continuación, se presentan diversos escritos a mano para ver cuáles son algunos de los problemas que se evidenciaron con la escritura en papel.

La caligrafía, puede incidir positiva o negativamente en el evaluador, tal como se aprecia en la Figura 5.2. En este caso, quizás la falta de familiarización con el alfabeto grecolatino haría suponer a primera vista que el texto fue escrito por un niño, lo cual puede incidir negativamente en la manera en la cual el evaluador percibe el nivel del sustentante tan sólo por su caligrafía. Aunque efectivamente el texto era muy básico, el evaluador ya tiene una primera impresión negativa de un texto.

Figura 5.2: Ejemplo de texto donde la caligrafía incide negativamente de forma visual

Estas dos gráficas habla sobre relación de nivel educativo de los padres y sus hijas.

La gráfica de los padres indica que no han ido a escuela 62.2%. Y que ban a la universidad no más cuenta 2.6%. Y que ban a posgrado es 0.2%. Esta gráfica nos enseñan que mayoría de generación de los padres no tuvieron la oportunidad de estudiar suficientemente. (Sería los padres que no tuvieron tener oportunidad de estudiar suficientemente.)

Y el otro lado, la gráfica de las hijas expresa que ellas tuvieron más la oportunidad para ir a la escuela que los padres. Quizás, razon es que la época esta abansando poco a poco para estar dentro de orden. Pero tenemos fijarse que toda vida hay los hijas que estan fuera de educación. Y notenemos que olvidarnos hasta que desaparese las gentes "sin estudios". Para futuro de los niños.

Transcripción literal del texto:

Estas dos gráficas habla sobre relación de nivel educativo de los padres y sus hijas. La gráfica de los padres indica que no han ido a escuela 62.2%. Y que ban a la universidad no más cuenta 2.6%. Y que ban a posgrado es 0.2%. Esta gráfica nos enseñan que mayoría de generación de los padres no tuvieron la oportunidad de estudiar suficientemente. Y el otro lado, la gráfica de las hijas expresa que ellas tuvieron más la oportunidad para ir a la escuela que los padres. Quizás, razon es que la época esta abansando poco a poco para estar dentro de orden. Pero tenemos fijarse que todavía hay las hijas que estan fuera de educación. Y notenemos que olvidarnos hasta que desaparese las gentes "sin estudios". Para futuro de los niños.

(131 palabras).

El tamaño de la letra influye en la percepción visual del evaluador, puesto que parece mucho más extenso de lo que en realidad es, tal como se muestra en la Figura 5.3. En el caso opuesto, cuando se revisaron los textos transcritos en computadora, éstos se percibían como demasiados pequeños, dado el contraste que se notaba entre los textos transcritos a mano versus en computadora. El siguiente texto fue escrito en la carátula completa de una hoja (por

razones de formato se ajustó al que aquí se presenta), por lo cual parecía mucho más extenso en papel.

Figura 5.3: Los textos en computadora se perciben de longitud menor en comparación con los escritos a mano

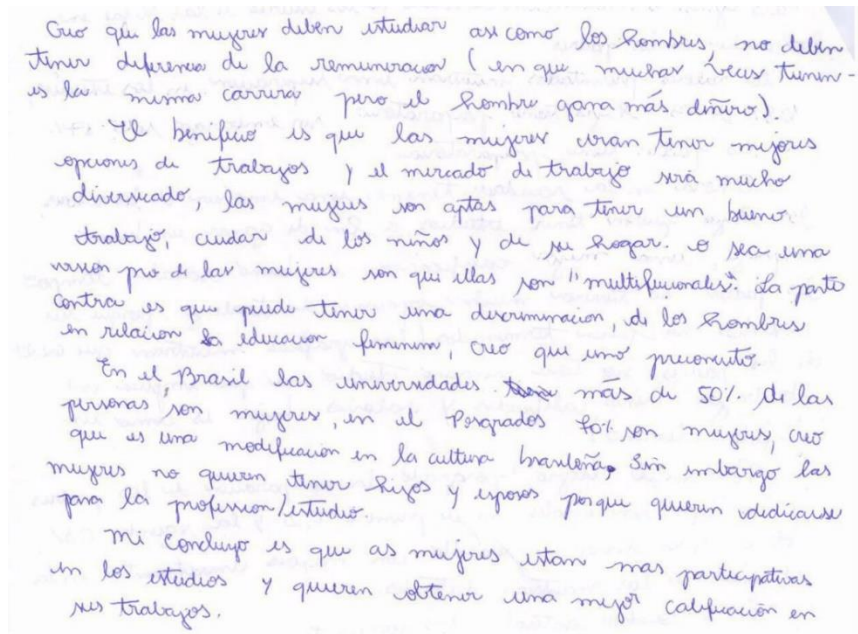
El documento se compone de dos gráficas que muestran los estudios que hicieron los padres y los que están haciendo sus hijas. En la primera gráfica podemos ver que el 62,2% de los padres no hicieron estudios aunque en la segunda gráfica solo el 26,3% de las hijas se quedan sin estudios. Si nos interesamos a las licenciaturas, el 6,4% de las hijas estudiaron en contra a los padres que representaron el 2,6%. En cuanto al posgrado, no hay una gran diferencia de porcentaje porque los padres representaron un 0,2% y las hijas un 0,3%. Sin embargo, la preparatoria contaba con 10,8% de hijas y sólo 1,7% de padres. Además, vemos que las hijas estaban inscritas en el secundaria y en la primaria y no era el caso de sus padres. Gracias a estas dos gráficas, vemos que las hijas, o sea, los niños de hoy estudian más que sus padres. El documento pone de

Transcripción del texto:

El documento se compone de dos gráficas que muestran los estudios que hicieron los padres y los que están haciendo sus hijas. En la primera gráfica podemos ver que el 62,2% de los padres no hicieron estudios aunque en la segunda gráfica sólo el 26,3% de las hijas se quedan sin estudios. Si nos interesamos a las licenciaturas, el 6,4% de las hijas estudiaron en contra a los padres que representaron el 2,6%. En cuanto al posgrado, no hay una gran diferencia de porcentaje porque los padres representaron un 0,2% y las hijas un 0,3%. Sin embargo, la preparatoria contaba con 10,8% de hijas y sólo 1,7% de padres. Además, vemos que las hijas estaban inscritas en el secundaria y en la primaria y no era el caso de sus padres. Gracias a estas dos gráficas, vemos que las hijas, o sea, los niños de hoy estudian más que sus padres. El documento pone de...

(181 palabras).

Finalmente, había textos que resultaban difíciles de leer, con lo cual la lectura se veía interrumpida y en ocasiones el evaluador debía interpretar mediante el contexto las palabras o paquetes léxicos que no comprendía. La Figura 5.4 presenta un ejemplo de este tipo. Debido a las problemáticas evidenciadas en las tareas propuesta para el pilotaje, se decidió llevar a cabo modificaciones para la primera aplicación oficial del examen. En primer lugar, con la implementación de la producción de textos dentro de la plataforma del examen se cambió el modo de escritura y los estudiantes escribieron en computadora. Esto incrementó considerablemente el número de palabras que utilizaron, sobre todo para los ensayos, en los cuales hubo sustentantes que incluso escribieron más de 700 palabras en un texto en el cual se solicitó un mínimo de 300. Esto da cuenta importante de la velocidad que implica no sólo la escritura sino también la edición de textos que son escritos en computadora. Vale la pena recordar que los estudiantes ya estaban registrados en algún posgrado de la UNAM, con lo cual se puede evidenciar que se encontraban familiarizados con el teclado en español.

Figura 5.4: La mala caligrafía dificulta la lectura de los textos

Con base en esta experiencia, se hicieron modificaciones en las tareas utilizadas. En primer lugar, se hizo una independencia de tareas; es decir, las dos abordaban temas completamente diferentes con el fin de explotar al máximo la riqueza léxica de los examinados. También se redujo la información proporcionada en la consigna, de tal suerte que no pudieran copiar información proveniente de ésta. En cuanto a la nueva propuesta de la gráfica, el único problema que se presentó fue con respecto a su interpretación, dado que 8 de los 31 estudiantes malinterpretaron la gráfica pensando que ésta mostraba que entre mayor era el número de hijos de mujeres mexicanas, éstos contaban con menor escolaridad, cuando en realidad la gráfica mostraba que las mujeres con mayor logro educativo tendían a concebir menor número de hijos (ver Anexo 2.5).

En el caso de la segunda tarea, se buscó un tema que pudiera ser de interés general para los estudiantes y en el cual se pudieran establecer argumentos a favor o en contra de una postura. En este caso, cabe mencionar que la mayoría de los estudiantes mostraron interés en la tarea propuesta y no se presentaron problemas en incluir contraargumentos en cuanto a los sistemas de salud públicos o privados. En general, la única problemática que se evidenció con la temática del ensayo fue que dos estudiantes reaccionaron emocionalmente ante el tema. El tema viró de un tono argumentativo a uno narrativo respecto de las malas experiencias médicas que habían vivido con el sistema de salud público en México, cuando en realidad no se les pedía

que narraran experiencias personales con el sistema de salud, sino que argumentaran en favor de uno de ellos.

5.4.2. LA CALIDAD DEL PROCESO DE EVALUACIÓN MEDIANTE MFRM

Debido a que el análisis estadístico de MFRM se emplea tanto para validar las distintas facetas: la tarea, la rúbrica y los evaluadores, en este apartado de los resultados se incorporan los resultados estadísticos de éstas tres. Esta sección de los resultados se encuentra dividida en dos fases: pilotaje con la primera versión de la rúbrica (ver Anexo 2.2) y evaluación con la segunda versión de la rúbrica (ver Anexo 2.6). De igual manera, como se mencionó en el apartado anterior, durante la fase del pilotaje, cada uno de los 37 estudiantes fue evaluado por dos de los cinco correctores (ver Anexo 2.3). Posteriormente, con base en la información estadística y en la experiencia de los evaluadores en el uso de esta rúbrica preliminar, se llevaron a cabo modificaciones que condujeron a la modificación de la rúbrica. En la segunda fase de evaluación, con la segunda versión de la rúbrica, 100 estudiantes (incluyendo la primera muestra de 37) fueron evaluados por 6 correctores en al menos una de las dos tareas (ver Anexo 2.7).

5.4.2.1. PILOTAJE: EVALUACIÓN CON LA PRIMERA VERSIÓN DE LA RÚBRICA

Para poder llevar a cabo el análisis estadístico de MFRM con la primera versión de la rúbrica, se tuvo que ajustar el puntaje de cada uno de los criterios con el fin de hacerlas comparables entre sí. Por ejemplo, en esta primera versión de la rúbrica, había criterios binarios (0–1 punto), criterios con niveles de 0 a 3 puntos y criterios con niveles de 1 a 3 puntos. Este desfase entre los criterios dificultó que se pudieran comparar cada uno de ellos, pues en realidad estadísticamente no eran comparables a través de MFRM, de tal suerte que tuvo que llevarse a cabo un proceso de estandarización de los criterios de modo que se evaluaran en parámetros del 1 al 3 cada uno de ellos. Solamente el desarrollo de ideas contaba con criterios binarios, de modo que se sacó el promedio obtenido en los diferentes criterios (cuatro para la tarea 1 y once para la tarea 2) y este número se multiplicó por tres, para poder estandarizar el criterio a 3 puntos.

El programa de FACETS arroja diversas tablas que sirven para poder interpretar el funcionamiento de cada una de las facetas, así como de la habilidad de los examinados. La Tabla 5.1 proporciona el resumen del análisis de todos los datos. En primer lugar, aparecen las categorías (*cat*); es decir, el promedio de las observaciones que se llevaron a cabo en total. Posteriormente, el puntaje que se obtuvo (*score*) así como el puntaje que se esperaba (*exp*). Más adelante, se observan los residuales (*resd*), que equivalen al puntaje obtenido menos el puntaje

esperado. Por último, se observan los residuales estandarizados (*StRes*). Lo óptimo es que la media de los residuales y los residuales estandarizados de los datos sean prácticamente iguales a cero, pues esperamos que las evaluaciones llevadas a cabo sean similares a las predichas por el modelo. Sin embargo, en el caso de la desviación estándar de los valores de la población o de la muestra, se espera que los residuales estandarizados sean semejantes a 1.0, pues esto indica que se trata de una distribución uniforme (Linacre, 2012b). En el extremo derecho de la Tabla 5.1, se aprecian estas dos categorías que denotan la diferencia entre una población y una muestra (*population vs sample*).

Tabla 5.1: Resumen de medición de datos

Cat	Score	Exp.	Resd	StRes	
2.26	2.26	2.26	.00	.00	Mean (Count: 1500)
.69	.69	.43	.54	1.00	S.D. (Population)
.69	.69	.43	.54	1.00	S.D. (Sample)

Data log-likelihood chi-square = 2291.7087
 Approximate degrees of freedom = 1448
 Chi-square significance prob. = .0000

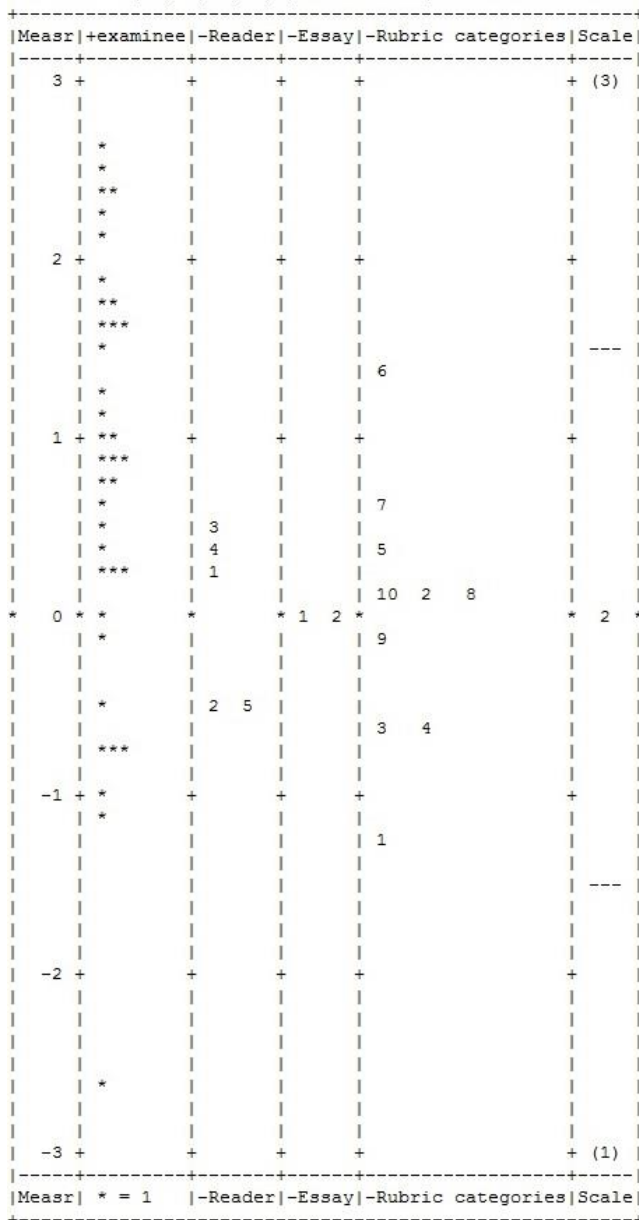
El mapa de variabilidad (también conocido como mapa de Wright) de la Figura 5.5 nos muestra las distintas facetas que se consideraron en el estudio: el candidato, el evaluador, la tarea y las categorías de la rúbrica. En la primera columna aparece la medida, misma que se encuentra determinada en lógitos. Esta medida especial que genera el programa FACETS establece un cero convencional como el punto medio de las facetas en cuestión. Linacre (2012b, p. 4) menciona que típicamente estas medidas en lógitos se encuentran en un rango de -5 a 5 lógitos. Esta medida parte de un cero establecido de manera convencional –punto medio de dificultad– y puede incrementar de forma positiva o negativa. Generalmente sólo la faceta 1 es positiva, lo cual indica que entre más se aleja del cero de forma positiva, mayor es la habilidad de escritura del sustentante¹². Consecuentemente, esperamos que los estudiantes promedio se encuentren en el cero, mientras que aquellos con mayores habilidades por arriba de éste. En el caso de las pruebas de escritura en los exámenes de lengua, el examinado se considera como la única faceta positiva, mientras que los examinadores, la tarea y los criterios de la rúbrica representarían las facetas negativas¹³; es decir, entre más se alejan del cero de forma positiva,

¹² Consúltese Linacre (2012, p. 15), tutorial de Facets 1, para ver los antecedentes de la definición de medición en las ciencias sociales.

¹³ Convención adoptada para la educación (Linacre, 2012, p. 26).

significa que el evaluador es más severo y la tarea y las categorías de la rúbrica más difíciles. En cada una de las facetas se encuentran los elementos. Éstos hacen referencia a un examinado en particular, un juez que lo evalúa, una tarea específica o un criterio de la rúbrica que recibe un puntaje. Entonces, idealmente lo que se busca es que cualesquiera de las facetas negativas (evaluador, tarea y categorías de la rúbrica) se encuentren cerca del cero (de lo contrario significa que éstas están produciendo variabilidad indeseada en los resultados de los examinados).

Figura 5.5: Mapa de variabilidad del pilotaje: primera versión de la rúbrica



Nomenclatura

Primera columna

Measr= medición, dada en lógitos.

Segunda columna

Examinee= examinados
* indica un candidato

Tercera columna

Reader= evaluador

Cuarta columna

Essay= texto
1 – descripción de gráficas
2 – redacción de ensayo argumentativo

Quinta columna

Rubric categories = categorías de la rúbrica
1 – Registro académico
2 – Desarrollo de ideas
3 – Cohesión
4 – Coherencia
5 – Léxico
6 – Precisión gramatical
7 – Ortografía
8 – Complejidad sintáctica
9 – Puntuación
10 – Competencia estratégica

Sexta columna

Niveles de la rúbrica
(1 a 3 puntos)

Finalmente, en la última columna aparecen los valores contemplados para la evaluación de la rúbrica: 1, 2 y 3. La línea punteada entre estos rangos denota el nivel umbral en el cual un candidato “x” habría tenido la misma posibilidad de haber caído en un rango o en otro. En este caso, estos niveles no son decisivos sobre el desempeño del candidato, pero podría haber casos en los que los estudiantes se encontraran dentro de este rango umbral y, por lo tanto, sería importante determinar si se encuentran en un nivel u otro.

Como se puede apreciar en el mapa de variabilidad del pilotaje (Figura 5.5), lo primero que destaca de la columna 2 (examinado) es que los estudiantes cuentan con una gran habilidad para escribir (la gran mayoría –cada examinado se encuentra representado por un *– se encuentra arriba del cero). Esto se debe a que las muestras corresponden a dos tipos de poblaciones que participaron en este primer período: 19 alumnos cursando estudios de posgrado y 18 alumnos de intercambio académico a nivel licenciatura –la gran mayoría incluso ya con certificados probatorios de su nivel de español igual o superior al B2–. En la segunda columna se aprecian los evaluadores. Se puede observar como los evaluadores 2 y 5 son los más indulgentes, mientras que los 1, 3 y 4, más severos. Sin embargo, ninguno se encuentra en el centro. En la tercera columna se observa que las tareas 1 (descripción y contraste de gráficas) y 2 (redacción de un ensayo de tipo argumentativo) representan la misma dificultad. De la cuarta columna sobresalen dos categorías: la 6 (gramática), que se evalúa con mayor dureza, y la 1 (registro académico), que se evalúa con mayor indulgencia. El resto de las categorías se encuentran más o menos dentro de los parámetros aceptables.

Si bien es cierto que el mapa de variabilidad es de gran utilidad para forjarse una rápida imagen de qué tan fiable es el proceso de evaluación, éste no proporciona un reporte detallado de la calidad de cada una de las facetas. Los derivados de las Tablas 5.2 sí arrojan dichos reportes. Aquí es donde se encuentran los datos más importantes del análisis de calidad del instrumento de evaluación y de los evaluadores. El número de tablas que arroja el programa depende del número de facetas con que se cuenta. En este caso, hay 4; por lo que la tabla 5.2a. hace referencia a los examinados, la 5.2b. a los evaluadores, la 5.2c. a las tareas y la 5.2d. a las categorías de la rúbrica. En éstas encontramos los estadísticos esenciales de cada elemento en cada faceta. En el extremo izquierdo, en la primera columna, aparecen cuatro subdivisiones que hacen referencia a: el puntaje total, el total de observaciones, el promedio observado y el puntaje justo. Sin lugar a dudas, una de las grandes ventajas de este programa es que establece un puntaje justo para el candidato, dependiendo de los ajustes que establece el modelo con base

en quien lo evaluó. En la segunda columna de izquierda a derecha, encontramos dos subdivisiones: la medición y el error estándar del modelo. La medición (*measure*) se encuentra en lógitos. En el caso de los examinados, la medición nos permite saber cuál es su habilidad, mientras que en las demás facetas, la medición establece qué tan severo/indulgente es un evaluador o que tan difícil/fácil es una tarea o la categoría de una rúbrica. Recordemos que estas facetas son negativas, entonces, cuando los elementos presentan mediciones positivas, significa que el evaluador es más severo o la tarea o la categoría de la rúbrica más difícil.

En el caso del error estándar (*S.E.*), éste se encuentra en rangos del 0 al 1. Idealmente, este error estándar debería ser menor a 0.05 –lo cual equivale a un margen de error del 5%–. El error estándar nos indica la precisión con la cual estamos calculando la información. Naturalmente entre más observaciones tengamos por parte de un evaluador, mayor precisión tendremos en la manera mediante la cual evalúa. Imaginemos por un momento que un tirador de arco lanza una flecha 100 veces. El lugar en donde mayoritariamente haya caído la flecha nos dará una mayor precisión de su habilidad para el tiro con arco; sin embargo, si sólo lanza la flecha 10 veces, entonces no tendremos tanta precisión en determinar su habilidad. Consecuentemente, entre más observaciones tengamos, menor será el error estándar y viceversa.

En la columna central figuran los índices de ajuste. Estos índices son los que nos sirven para controlar la calidad de cada una de las facetas en cuestión (i.e., el examinado, la tarea, el evaluador y las categorías de la rúbrica). Estos índices de ajuste se denominan *Infit MnSq*¹⁴ y *Outfit MnSq*. El índice de ajuste muestra si los puntajes observados se parecen o no a los puntajes predichos por el modelo. Estas predicciones se llevan a cabo con base en el nivel del sustentante, la severidad del evaluador y la dificultad de la tarea. A la diferencia entre los puntajes observados de los esperados se le conoce como residuales estandarizados, e idealmente deberían ser iguales a 0. Cuando estos residuales se elevan al cuadrado y éstos se suman a lo largo de las diferentes facetas y elementos de cada una de ellas, se obtienen los índices de ajuste para el modelo (Eckes, 2009a, Linacre, 2012b). El índice de ajuste *Infit (inlier fit)* es la media ponderada de los cuadrados de las diferencias estandarizadas y el *Outfit (outlier fit)* es la media no ponderada de los cuadrados de las diferencias estandarizadas (Linacre, 2013, p. 191).

¹⁴ El valor MnSq (mean-square) hace referencia al estadístico de ajuste de la Chi-cuadrada entre sus grados de libertad. Consúltese Linacre (2012b, p. 39) para mayor información sobre la Chi-cuadrada y los grados de libertad.

El índice *Outfit* es particularmente sensible a los resultados atípicos, mientras que el *Infit* es sensible a los patrones de respuestas específicas. Dicho en otras palabras, el índice *Outfit* es sensible a las respuestas con valores extremos (p.ej., que un estudiante conteste correctamente ítems difíciles y no los intermedios), mientras que el *Infit* es sensible al patrón de respuestas (i.e., que un estudiante presente inconsistencias en su patrón de respuestas). Los valores del *Outfit* e *Infit* pueden oscilar en un rango de 0 a ∞ , aunque el valor esperado es 1. Los valores mayores a 1.2 indican un desajuste mayor a lo esperado, y si este valor llega a 2 se considera como un severo desajuste (las observaciones serían demasiado impredecibles), lo cual significa que las observaciones resultan aleatorias y poco sistemáticas (Linacre, 2012b). Finalmente, los valores menores a 1 se pueden interpretar como sobreajustes (las observaciones son demasiado predecibles).

El autor hace una analogía con la música y menciona que si nuestros datos se encuentran entre 0.5 y 1.5 (rango productivo para la medición), es como si se escuchara música placentera y con los niveles óptimos de calidad y de volumen. Valores entre 1.5 a 2.0 (no productivos para la medición, pero tampoco degradantes para ésta), sería como escuchar música con ruido de fondo perceptible, pero que no distorsiona la música. Valores menores a 0.5 (poco productivos para la medición, pero no degradantes de ésta) equivaldrían a escuchar música muy tenue. Finalmente, los valores mayores a 2.0 (distorsionan o degradan el sistema de medición) se equipararían a escuchar música en donde el ruido externo es tal que distorsiona la música (Linacre, 2012b, p. 11). Esta información se resume en el Cuadro 5.4. En el caso de los exámenes de alto impacto o administrados a gran escala, Linacre (2012b, 2013) sugiere que el rango aceptable oscile entre 0.8 a 1.2. Por último, Linacre aconseja que primero se revisen los índices arriba de 1.2, pues los índices menores a 0.8 no causan realmente un problema en la medición. Asimismo, el autor recomienda que se preste mayor atención a los valores del *Outfit*, aunque también pueden reportarse ambos (2012b, p. 25).

Cuadro 5.4: Interpretación de los índices de ajuste (Linacre, 2012b, p. 11)

(Traducido por el investigador)

Media del cuadrado	Interpretación
> 2.0	Distorsión o degradación del Sistema de medición.
1.5 – 2.0	No productivo para la medición, pero no degrada la medición.
0.5 – 1.5	Productivo para la medición.
< 0.5	Poco productivo para la medición, pero no la degrada.

A diferencia del error estándar que nos indica la precisión, los índices de ajuste nos muestran que tan exacto es un evaluador. Por ejemplo, regresando al ejemplo del tirador con arco, un participante puede colocar 90 por ciento de veces la flecha en la banda azul (digamos que el centro es el color amarillo). Eso indica que es muy preciso, pero no muy buen tirador, pues no es muy certero; es decir, nunca llega al centro. Sin embargo, puede ser que encontremos un tirador que el 90 por ciento de las veces coloca la flecha en la banda amarilla, directo en el centro. Esto quiere decir que el tirador es muy preciso, pero también extremadamente exacto. Por esta razón, las rúbricas analíticas permiten también una mejor precisión para determinar la habilidad de un sustentante; puesto que, en el caso de la PE, el candidato es observado en siete categorías, en dos tareas, por dos evaluadores, lo cual significa que es observado y puntuado 28 veces. Este dato será mucho más preciso que si se emplea una rúbrica holística en la cual el candidato sería solamente observado y evaluado 4 veces (dos evaluadores y dos tareas). Viéndolo de manera análoga con otras habilidades, es más preciso un examen de comprensión de lectura que contiene 28 preguntas, que uno que contiene solamente 4, puesto que podemos observar en mayores ocasiones la habilidad del examinado de forma detallada.

En la penúltima columna de izquierda a derecha encontramos las correlaciones. En la primera tabla figuran aquellas correlaciones observadas, también denominadas como puntos biserials (*point-biserial*) o medida de puntos de correlación (*PtMea*). FACETS espera que los evaluadores demuestren una independencia en las evaluaciones, por lo cual, dependiendo de diversos criterios –tales como los niveles contemplados en la rúbrica y el número de veces que observó el evaluador– el modelo establece los puntos de correlación esperados. Si las correlaciones son casi perfectas, entonces esto indica que los evaluadores se están comportando como máquinas y no como evaluadores independientes.

“Facets models the raters to be “independent experts”. These would produce an “exact agreement” percent, which is the same or slightly higher than the “expected agreement” percent. But many raters are trained to behave like “rating machines”. Agreement is encouraged among the raters, and disagreements are penalized. For these raters we expect the “exact agreement” percent to be much higher than the “expected agreement” percent. When the “exact agreement” approaches 100%, the raters are behaving the same way as optical scanners do for “bubble sheets”. The raters have become part of the data-collection mechanism, they are no longer a facet of the measurement situation (Linacre, 2012b, p.26).

Por lo tanto, el valor de *PtMea* deberá acercarse lo más posible al punto de correlación esperado (*PtExp*), y no al 100%. No obstante, Linacre (2012b, 2013) afirma que cuando estos valores son inferiores a 0.3, no hay consistencia en la evaluación. Finalmente, en la última

columna de la derecha encontramos la faceta que estamos analizando: el examinado, el evaluador, la tarea o las categorías de la rúbrica¹⁵.

La Tabla 5.2a. nos muestra el reporte de medición de los estudiantes. Un dato importante de esta tabla es que nos muestra el puntaje justo que merece un estudiante sin importar quién lo evaluó. FACETS establece este ajuste automáticamente. Los alumnos se encuentran jerarquizados de menor a mayor habilidad. En la tabla figura en primer lugar el alumno 12 con menor habilidad (-2.67 lógitos equivalente a un puntaje de 1.25 y un puntaje justo de 1.21, tomando en cuenta que el máximo de la rúbrica es 3 y el mínimo 1). En la primera columna podemos ver que este alumno fue observado 40 veces (2 tareas, 2 evaluadores, 10 categorías de la rúbrica) y recibió un puntaje acumulado de 50 puntos. En el caso de las demás columnas del reporte de los estudiantes, se revisarán con mayor detalle en el siguiente apartado con los resultados arrojados por la segunda versión de la rúbrica.

La Tabla 5.2b. nos muestra el reporte de medición de los evaluadores. Esta tabla es la más importante para poder evidenciar la consistencia y confiabilidad de los evaluadores. En la última columna encontramos al evaluador. En la primera columna se puede observar cuántas observaciones llevó a cabo y el total acumulativo que obtuvo y el promedio. Esta tabla también establece el puntaje medio justo para cada uno de los evaluadores. En la segunda columna se aprecia la medición en lógitos y también el error estándar de cada evaluador. Lo primero que se percibe es que el evaluador más indulgente es el 2 (-.55 lógitos) y el más severo es el 3 (.56). La suma de ambos nos da un rango de severidad de 1.12, el cual es aceptable¹⁶. Sin embargo, en el error estandar podemos ver valores menores a 0.5 en solamente dos evaluadores (1 y 2), mientras que el 5, por ejemplo, presenta un *S.E.* de 0.31. Esto se debe a que fueron muy pocos

¹⁵ Consúltense Linacre (2012b y 2013) para información detallada sobre otros aspectos de las Tablas 5.2 (*Zstd* que son las medias cuadradas estandarizadas como estadístico *z* (*z-statistics*), el valor de *RMSE* (*root mean-square error*) que hacen referencia al promedio estadístico del error estándar de las mediciones, el *Adj (True) SD* que significa “la verdadera desviación estándar” después de ser ajustada con el error de medida, la separación (*separation*) que indica cuántos estratos de medición podrían ser estadísticamente distinguibles entre las mediciones y el índice de confiabilidad (*reliability*) que no hace referencia a la confiabilidad del jueceo de los evaluadores, sino al grado de consistencia del modelo).

¹⁶ Eckes (2012, p. 280) sugiere que el rango de severidad de los evaluadores no exceda la cuarta parte del rango de habilidad observado por parte de los examinados. En la Tabla 5.2a. podemos ver que el rango menor fue de -2.67 lógitos y el más alto de 2.59, lo cual nos da un rango de 5.26 lógitos (habilidad de los estudiantes), cifra que si dividimos entre 4 nos da un total de 1.31 lógitos de rango aceptable entre elementos de una faceta (i.e., tarea, evaluador y categorías de la rúbrica).

textos los que evaluó y lo que demuestra es que la seguridad con la cual podemos medir su capacidad de evaluar no es muy precisa.

Tabla 5.2a: Reporte de medición de los estudiantes con la primera versión de la rúbrica

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Nu examinee
50	40	1.25	1.21	-2.67	.37	1.26	1.0	1.34	.9	.70	.00	.35	12 12
67	40	1.67	1.60	-1.15	.28	1.01	.1	.97	.0	1.00	.38	.41	29 29
69	40	1.73	1.65	-.99	.28	.69	-1.7	.69	-1.7	1.42	.37	.42	25 25
72	40	1.80	1.73	-.77	.27	1.05	.2	1.03	.2	1.02	.71	.42	23 23
70	40	1.75	1.74	-.75	.28	.83	-.8	.89	-.4	1.11	.01	.46	14 14
73	40	1.83	1.76	-.69	.27	.57	-2.4	.59	-2.4	1.58	.47	.42	35 35
68	40	1.70	1.84	-.45	.28	.61	-2.2	.61	-2.3	1.60	.65	.38	17 17
80	40	2.00	1.94	-.18	.27	.81	-.9	.81	-.9	1.28	.58	.43	38 38
82	40	2.05	1.99	-.03	.27	1.75	3.1	1.79	3.3	-.14	-.22	.43	26 26
85	40	2.13	2.07	-.19	.27	1.32	1.5	1.32	1.5	.51	.11	.43	27 27
85	40	2.13	2.07	-.19	.27	1.10	.5	1.08	.4	.90	.55	.43	39 39
86	40	2.15	2.09	.27	.27	.81	-.9	.81	-.9	1.30	.64	.42	30 30
80	40	2.00	2.15	.44	.27	1.24	1.1	1.23	1.1	.71	.51	.39	16 16
88	40	2.20	2.17	.47	.27	.95	-.1	.95	-.1	1.05	.37	.44	11 11
91	40	2.28	2.23	.64	.28	.92	-.3	.93	-.2	1.11	.50	.42	24 24
90	40	2.25	2.27	.77	.28	.84	-.8	.83	-.8	1.28	.65	.46	19 19
93	40	2.33	2.28	.80	.28	1.03	.2	1.03	.2	.99	.57	.42	5 5
86	40	2.15	2.29	.82	.27	.76	-1.2	.76	-1.2	1.32	.32	.39	3 3
93	40	2.33	2.30	.86	.28	1.02	.1	.98	.0	1.01	.49	.43	13 13
89	40	2.22	2.32	.92	.27	1.55	2.4	1.52	2.3	.19	.04	.38	2 2
95	40	2.38	2.33	.96	.28	.86	-.6	.83	-.8	1.28	.70	.41	6 6
139	60	2.32	2.36	1.02	.23	.85	-.9	.82	-1.1	1.30	.70	.43	7 7
94	40	2.35	2.37	1.07	.28	.69	-1.6	.68	-1.7	1.43	.50	.45	10 10
99	40	2.47	2.44	1.29	.29	1.09	.5	1.19	.9	.81	.29	.40	34 34
101	40	2.53	2.50	1.47	.30	.75	-1.3	.69	-1.5	1.42	.71	.39	28 28
102	40	2.55	2.53	1.56	.31	1.39	1.8	1.52	2.1	.40	-.08	.39	33 33
98	40	2.45	2.55	1.62	.29	1.37	1.7	1.36	1.7	.45	.13	.36	4 4
101	40	2.53	2.56	1.69	.30	1.10	.5	1.00	.0	.91	.41	.43	18 18
104	40	2.60	2.58	1.76	.32	1.10	.5	1.04	.2	.89	.28	.38	31 31
102	40	2.55	2.59	1.77	.31	.69	-1.6	.65	-1.6	1.40	.55	.42	15 15
105	40	2.63	2.61	1.86	.32	.71	-1.4	.67	-1.4	1.35	.54	.37	36 36
105	40	2.63	2.67	2.07	.33	1.03	.1	.96	.0	.95	.25	.40	8 8
108	40	2.70	2.69	2.19	.35	1.37	1.5	1.54	1.7	.70	.27	.35	1 1
109	40	2.72	2.72	2.32	.36	1.00	.0	1.14	.5	.93	.19	.34	32 32
109	40	2.72	2.74	2.39	.36	.69	-1.3	.70	-.9	1.28	.56	.35	9 9
110	40	2.75	2.75	2.45	.37	.94	-.1	.97	.0	1.01	.28	.33	22 22
111	40	2.78	2.78	2.59	.38	1.67	2.1	1.39	1.0	.60	.15	.32	37 37
91.6	40.5	2.26	2.26	.78	.30	1.01	.0	1.01	-.1		.38		Mean (Count: 37)
16.4	3.2	.36	.37	1.17	.03	.29	1.3	.30	1.3		.24		S.D. (Population)
16.7	3.3	.37	.38	1.19	.04	.29	1.4	.30	1.4		.24		S.D. (Sample)
Model, Populn: RMSE .30 Adj (True) S.D. 1.13 Separation 3.79 Strata 5.39 Reliability .93													
Model, Sample: RMSE .30 Adj (True) S.D. 1.15 Separation 3.84 Strata 5.46 Reliability .94													
Model, Fixed (all same) chi-square: 496.3 d.f.: 36 significance (probability): .00													
Model, Random (normal) chi-square: 33.6 d.f.: 35 significance (probability): .54													

Eckes (2009a) y Linacre (2012) mencionan que cuando hay más muestras observadas, la precisión del modelo es mayor y, por lo tanto, el error estándar disminuye. En la columna central está la información más importante, relacionada con los índices de calidad de la medición. En realidad, todos los evaluadores se encuentran dentro de un rango muy aceptable tanto en valores *Outfit* como *Infit* (0.8 a 1.2), lo cual indica que las evaluaciones son consistentes y también confiables. Esto se puede corroborar en la columna de las correlaciones, donde se aprecia que

las correlaciones observadas de todos los evaluadores (*PtMea*) se encuentran por encima de las esperadas (*PtExp*).

Tabla 5.2b: Reporte de medición de los evaluadores con la primera versión de la rúbrica

Total Score	Total Count	Obsvd Average	Fair(M) Average	Model Measure	Infit S.E.	Outfit MnSq ZStd	Estim. Discrm	Correlation PtMea PtExp	Exact Obs %	Agree. Exp %	N Reader	
1437	600	2.39	2.46	-.55	.08	.87 -2.3	1.16	.67 .62	56.1	50.4	2 2	
102	40	2.55	2.45	-.52	.31	.96 -1.1	1.11	.54 .40	52.5	46.1	5 5	
1102	500	2.20	2.20	.20	.08	1.18 3.0	1.22 3.5	.77	.57 .59	56.7	50.1	1 1
306	140	2.19	2.16	.32	.15	1.12 1.1	1.13 1.1	.81	.44 .53	51.4	49.3	4 4
442	220	2.01	2.08	.56	.12	.82 -2.2	.90 -1.0	1.15	.59 .62	54.2	48.3	3 3
677.8	300.0	2.27	2.27	.00	.15	.99 -.1	.99 .1	.56				Mean (Count: 5)
506.3	214.3	.19	.15	.45	.08	.14 2.1	.15 2.1	.07				S.D. (Population)
566.1	239.6	.21	.17	.51	.09	.16 2.3	.17 2.3	.08				S.D. (Sample)

Model, Populn: RMSE .17 Adj (True) S.D. .42 Separation 2.48 Strata 3.64 Reliability (not inter-rater) .86
 Model, Sample: RMSE .17 Adj (True) S.D. .48 Separation 2.81 Strata 4.09 Reliability (not inter-rater) .89
 Model, Fixed (all same) chi-square: 83.0 d.f.: 4 significance (probability): .00
 Model, Random (normal) chi-square: 3.6 d.f.: 3 significance (probability): .31
 Inter-Rater agreement opportunities: 780 Exact agreements: 433 = 55.5% Expected: 387.9 = 49.7%

En la Tabla 5.2c. se presenta el reporte de medición de las tareas que se emplearon. Tarea 1: descripción contraste de gráficas, tarea 2: ensayo argumentativo. Como se puede apreciar, la medición en lógitos de ambas es casi igual a cero, lo que indica que ambas tareas presentan prácticamente el mismo índice de dificultad. El error estándar es ligeramente superior a 0.05 y también los índices de calidad (*Outfit* e *Infit*) se encuentran prácticamente en 1, lo que indica que la medición es la óptima. Finalmente, Las correlaciones observadas son iguales a las esperadas.

Tabla 5.2c: Reporte de medición de las tareas con la primera versión de la rúbrica

Total Score	Total Count	Obsvd Average	Fair(M) Average	Model Measure	Infit S.E.	Outfit MnSq ZStd	Estim. Discrm	Correlation PtMea PtExp	N Essay	
1700	750	2.27	2.28	-.03	.07	1.01 .3	1.01 .2	.99 .62	2 2	
1689	750	2.25	2.26	.03	.07	.99 -.2	1.00 .0	1.02 .62	1 1	
1694.5	750.0	2.26	2.27	.00	.07	1.00 .0	1.01 .1	.62		Mean (Count: 2)
5.5	.0	.01	.01	.03	.00	.01 .3	.01 .1	.00		S.D. (Population)
7.8	.0	.01	.01	.04	.00	.02 .4	.01 .1	.00		S.D. (Sample)

Model, Populn: RMSE .07 Adj (True) S.D. .00 Separation .00 Strata .33 Reliability .00
 Model, Sample: RMSE .07 Adj (True) S.D. .00 Separation .00 Strata .33 Reliability .00
 Model, Fixed (all same) chi-square: .3 d.f.: 1 significance (probability): .60

En la Tabla 5.2d. se observa el reporte de medición de cada una de las categorías de la rúbrica. En la primera columna aparecen los totales, el número de observaciones (*total count*) y el promedio observado, así como el justo. Aquí se puede observar como la distancia entre la media de la categoría 1 (registro académico) y de la 6 (gramática) es muy grande. Esto se corrobora también en la segunda columna en la medición en lógitos (-1.25 para la adecuación y +1.34 para la gramática), lo cual indica un total de 2.59 lógitos de diferencia, valor inaceptable

(el rango calculado, siguiendo los parámetros sugeridos por Eckes, 2009a, era de 1.315). Es decir, se evalúa con demasiada dureza la gramática y con mucha indulgencia el registro académico. Adicionalmente, se observa en los índices de calidad (*Outfit* e *Infit*) un desajuste en la categoría 1 (registro académico) y la 10 (competencia estratégica). También las categorías 4 y 5 presentan valores ligeramente por debajo de lo que se espera (0.8). Adicionalmente, en las correlaciones también observamos ligeros desajustes. La categoría 1 (registro académico) y 2 (desarrollo de ideas) presentan correlaciones menores a las esperadas. Consecuentemente, debido a la inestabilidad en la calidad del instrumento de medición, aunado con las discusiones que se llevaron a cabo respecto del funcionamiento de la rúbrica –tras su uso durante el pilotaje– se llevaron a cabo enmiendas en ésta, por lo que se diseñó una segunda versión de la rúbrica (ver Anexo 2.6).

Tabla 5.2d: Reporte de medición de las categorías de la primera versión de la rúbrica

Total Score	Total Count	Obsvd Average	Fair (M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Nu Rubric categories
390	150	2.60	2.66	-1.25	.17	1.36	2.7	1.39	2.2	.66	.43	.54	1 1
367	150	2.45	2.48	-.63	.16	.92	-.7	.83	-1.4	1.17	.67	.57	3 3
367	150	2.45	2.48	-.63	.16	.60	-4.2	.65	-3.1	1.45	.68	.57	4 4
347	150	2.31	2.32	-.15	.15	1.17	1.5	1.14	1.2	.81	.57	.58	9 9
337	150	2.25	2.25	.07	.15	.92	-.7	1.05	.4	.94	.35	.58	2 2
336	150	2.24	2.24	.09	.15	.85	-1.4	.85	-1.4	1.22	.64	.58	8 8
332	150	2.21	2.21	.18	.15	1.38	3.3	1.35	2.9	.55	.58	.58	10 10
325	150	2.17	2.16	.33	.15	.68	-3.4	.71	-3.0	1.37	.57	.58	5 5
311	150	2.07	2.05	.63	.15	1.18	1.7	1.15	1.3	.78	.59	.58	7 7
277	150	1.85	1.80	1.34	.15	.98	-.1	.96	-.3	1.06	.63	.56	6 6
338.9	150.0	2.26	2.26	.00	.15	1.00	-.1	1.01	-.1		.57		Mean (Count: 10)
30.2	.0	.20	.23	.69	.01	.25	2.4	.24	2.0		.10		S.D. (Population)
31.8	.0	.21	.24	.73	.01	.26	2.5	.25	2.1		.10		S.D. (Sample)

Model, Populn: RMSE .15 Adj (True) S.D. .67 Separation 4.40 Strata 6.21 Reliability .95
 Model, Sample: RMSE .15 Adj (True) S.D. .71 Separation 4.65 Strata 6.54 Reliability .96
 Model, Fixed (all same) chi-square: 194.4 d.f.: 9 significance (probability): .00
 Model, Random (normal) chi-square: 8.6 d.f.: 8 significance (probability): .38

5.4.2.2. MODIFICACIÓN DE LA PRIMERA VERSIÓN DE LA RÚBRICA ANALÍTICA

Con base en los análisis de los resultados de la primera fase: pilotaje, se rediseñó la rúbrica siguiendo dos criterios: a) los análisis estadísticos: MFRM y b) la experiencia de los evaluadores al momento de evaluar mediante la rúbrica¹⁷.

¹⁷ Tras la primera fase de evaluación con la primera rúbrica, se llevaron a cabo 2 sesiones grupales de retroalimentación con los cinco evaluadores que participaron como jueces. Posteriormente, la responsable del examen y el investigador se reunieron en otras ocasiones para afinar la rúbrica con base en los análisis y en la retroalimentación de los evaluadores.

En primer lugar, se llevaron a cabo cambios de formato para homogenizar las puntuaciones otorgadas con base en los descriptores del nivel de logro de los sustentantes. Como se describió anteriormente, la rúbrica inicial incluía criterios que eran evaluados de forma binaria, otros más en un rango del 0 al 3 y otros cuantos del 1 al 3. No obstante, para poder llevar a cabo los procedimientos estadísticos de MFRM, se requería contar con valores que pudieran ser comparables entre sí. De tal suerte que el primer cambio que se implementó fue contar con rangos de 0 a 3 para cada uno de los rubros de la rúbrica. En segundo lugar, se identificaron aquellos componentes que no estaban funcionando bien bajo los dos criterios anteriormente descritos: análisis estadístico en MFRM y experiencia de los evaluadores al emplear la rúbrica. Primeramente, se abordará la categoría que se descartó de la rúbrica: la competencia estratégica y, posteriormente, se describirán aquellos que se reestructuraron con base en los resultados obtenidos.

La competencia estratégica presentó problemas moderados en el análisis de MFRM (índice de *Outfit* de 1.35). Sin embargo, el mayor conflicto con dicha categoría se debió a que, por una parte, los evaluadores expresaron dificultades al momento de emplear esta categoría, dado que resultaba difícil de interpretar y observar en los textos. Como se puede apreciar en el Cuadro 5.5, en la competencia estratégica se evalúan aspectos relacionados con la planeación y el monitoreo. Sin embargo, resultaba difícil evidenciar esto en escritos que no contaban con una hoja de planeación ni con muestras de reparación en el texto. Por otra parte, había examinados que sí presentaban dichas evidencias, pero no por ello mejoraba la calidad de sus textos, en comparación con quienes no lo hacían. Quizás los examinados con mayor habilidad cuentan con esta estrategia internalizada y automatizada, de tal suerte que no necesitan de un esquema de planeación y sus textos requieren una mínima edición. Además, se tenía ya contemplado que eventualmente los textos serían escritos en un procesador de textos sin corrector ortográfico, por lo cual incluir dicha categoría era innecesario. Por estas razones se excluyó esta categoría de la rúbrica.

Cuadro 5.5: Competencia estratégica (primera versión de la rúbrica)

Evidencia de la aplicación de alguna estrategia en alguna de las tareas.	Puntos	T1	T2	Total
Planeación y monitoreo (o un texto sin evidencia de estrategias, pero muy bien escrito)	3			/6
Planeación (esquemas, diagramas, lluvia de ideas).	2			
Monitoreo (reparación, reajuste de la tarea).	1			
Sin evidencia de estrategias en un texto fallido.	0			

Las categorías 8 (complejidad sintáctica) y 9 (puntuación) fueron modificadas y reubicadas en otras categorías que evalúan aspectos superiores de escritura: cohesión y coherencia. En el caso de la complejidad sintáctica (ver Cuadro 5.6), los evaluadores comentaron que resultaba difícil identificar y contabilizar las estructuras subordinadas. Además de que, en el caso de las gráficas, se podía hacer la descripción y contraste de manera eficiente sin recurrir abundantemente a la subordinación o formas variadas de nexos subordinantes.

Cuadro 5.6: Complejidad sintáctica (primera versión de la rúbrica)

	Puntos	T1	T2	Total
Uso de 5 o más nexos subordinantes distintos, sin errores (p. ej. mal empleo de nexo o problemas de correlación entre ellos). Se incluyen otros nexos además de <i>que</i> , <i>ya que</i> , <i>porque</i> y <i>aunque</i> .	3			/6
Uso de 4 nexos subordinantes distintos, sin error (o 5 nexos subordinantes distintos hasta con un error). Uso de los nexos más característicos (<i>que</i> , <i>ya que</i> , <i>porque</i> , <i>aunque</i>).	2			
Uso de 3 o menos nexos subordinantes distintos sin errores.	1			
Uso de 3 o menos nexos subordinantes distintos con errores.	0			

Con respecto a la puntuación (ver Cuadro 5.7), era claro que en muchas ocasiones, sobre todo en la descripción de gráficas, los examinados no requerían de ciertos signos exclamativos, interrogativos, paréntesis o incluso dos puntos para tener un texto muy coherente, lógico y bien estructurado con información precisa y clara. Por esta razón, además de pertenecer a una cuestión inferior de escritura, se decidió incorporar estos criterios también en la cohesión.

Cuadro 5.7: Puntuación (primera versión de la rúbrica)

	Puntos	T1	T2	Total
Sin errores de puntuación e inclusión de signos como <i>¡! ¿? ; : () — ... ""</i>	3			/6
Uso correcto del punto (.) y de la coma (,), pero con errores en el uso de signos como <i>¡! ¿? ; : () — ... ""</i>	2			
Uso correcto del punto (.) y de la coma (,)	1			
Errores de puntuación incluso en el uso del punto y coma.	0			

En relación con las demás categorías de la rúbrica, se hicieron modificaciones con base en el análisis de FACETS en combinación con los comentarios vertidos por los evaluadores en las sesiones de armonización de criterios de la rúbrica. El primer cambio importante fue el de la categoría 1 (registro académico). Los Cuadros 5.8 y 5.9 muestran un comparativo de la primera y segunda versión de la rúbrica. En FACETS, esta categoría fue la que presentó los niveles de *Outfit* más elevados (1.39), además de que los evaluadores reportaron problemas para poder adjudicar el nivel correspondiente, dado que no resultaba suficiente la descripción de

“convenciones formales académicas”. En efecto, la descripción de aspectos formales no resultaba suficiente para poder discriminar entre los alumnos bajos de los altos, incluso los estudiantes con habilidades de escritura limitadas obtuvieron un puntaje alto en este criterio.

Cuadro 5.8: Registro académico (primera versión de la rúbrica)

	Puntos	T1	T2	Total
Sin ninguna falla en el uso de las convenciones formales académicas : título, registro formal, tratamiento impersonal, objetividad (evitar anécdotas y frases como <i>yo creo que...</i>)	3			/6
Hasta 2 de las siguientes fallas de adecuación: ausencia de título (o utilización del mismo como introducción al primer párrafo), registro informal (uso de coloquialismos, barbarismos, palabras truncadas, etc.), inconsistencia en el tratamiento (cambio de persona gramatical), inserción de anécdotas personales y frases como <i>yo creo que...</i>	2			
Inadecuado (más de 2 fallas de adecuación).	1			

Cuadro 5.9: Registro académico (segunda versión de la rúbrica)

Adecuación: enfoque en el tema y uso de las convenciones formales.		
MARCA	PTOS	DESCRIPCIÓN
EXCEPCIONAL	3	El discurso logra plenamente su propósito comunicativo . Sin fallas en el uso de las convenciones formales académicas : registro formal, tratamiento impersonal, objetividad (evitar anécdotas y frases como <i>yo creo que...</i>). Se adecua al número de palabras solicitado (en un rango de 10% mayor o menor).
HÁBIL	2	El discurso logra su propósito comunicativo , pero presenta alguna falla en cuanto a: Inserción de anécdotas personales y frases como <i>yo creo que...</i> ; cita textual de las palabras utilizadas en la consigna. No se ajusta al número de palabras solicitado (en un rango de 20% mayor o menor).
EN DESARROLLO	1	El discurso NO logra su propósito comunicativo . Varias fallas de adecuación: Inserción de anécdotas personales y frases como <i>yo creo que...</i> ; cita textual de las palabras utilizadas en la consigna. No se ajusta al número de palabras (en un rango menor a 50%)
	0	El discurso NO logra su propósito comunicativo . Texto <i>fuera de las convenciones formales</i> : registro informal (uso de coloquialismos, barbarismos, palabras truncadas, etc.), inconsistencia en el tratamiento (cambio de persona gramatical). Texto demasiado corto para ser evaluado.

De hecho, también fue esta la categoría evaluada de forma más indulgente: 2.6 (promedio de la escala de la rúbrica de 0 a 3 puntos), equivalente a -1.25 lógitos (dato que ya por sí solo casi rebasa los límites establecidos como parámetros de indulgencia y severidad: 1.31 lógitos). Otro rubro que resultaba importante revisar, de acuerdo con los estadísticos de FACETS, fue la

gramática. Como se puede apreciar en el Cuadro 5.10, los criterios de evaluación de esta categoría eran demasiado severos y orientados hacia el error, mas no hacia el logro del sustentante. La puntuación media de la gramática fue de 1.85 (casi un punto de contraste con la adecuación en una escala del 1 al 3, lo cual era inadmisible), equivalente a 1.34 lógitos (dato que por sí solo rebasa el parámetro de indulgencia y severidad de los evaluadores). Como se puede observar, se describió en la rúbrica que solamente aquellos textos libres de errores recibirían la puntuación más alta. Sin embargo, dado que se trata de textos escritos en segunda lengua, siempre existirán pequeños errores e imprecisiones gramaticales, pese a que los examinados contaban con un amplio dominio en su habilidad de escritura. Por tal motivo, se decidió reestructurar la descripción de esta categoría (cfr. Cuadros 5.10 y 5.11).

Cuadro 5.10: Precisión gramatical (primera versión de la rúbrica)

	Puntos	T1	T2	Total
Sin errores en el uso de verbos, adverbios, pronombres, preposiciones, conjunciones, gerundios y locuciones.	3			/6
Hasta 3 errores en el uso de verbos, adverbios, pronombres, preposiciones, conjunciones, gerundios y locuciones.	2			
Más de 3 errores en el uso de verbos, adverbios, pronombres, preposiciones, conjunciones, gerundios y locuciones.	1			

Cuadro 5.11: Precisión gramatical (segunda versión de la rúbrica)

Precisión gramatical. Uso de las categorías gramaticales, modos y tiempos verbales (no incluye concordancia).		
MARCA	PTOS	DESCRIPCIÓN
EXCEPCIONAL	3	Uso preciso de categorías gramaticales. Prácticamente sin errores.
HÁBIL	2	Uso adecuado de las categorías gramaticales, con algunas imprecisiones .
EN DESARROLLO	1	Errores en el uso de las categorías gramaticales que evidencian la estructura sintáctica de otra lengua .
	0	Múltiples errores en el uso de las categorías gramaticales que impiden la comunicación del mensaje .

Por último, cabe mencionar que con base en las observaciones del responsable del examen y del investigador se hicieron otras enmiendas a las demás categorías de la rúbrica, pues daba la impresión que en primer término se estaban evaluado los errores cometidos por los sustentantes y no su desempeño en cada uno de las categorías. Por ende, la segunda versión de la rúbrica se reestructuró de tal manera que se evidenciara el logro del candidato en cada uno de los rubros. Estos cambios permitieron diseñar una nueva rúbrica más simplificada con mayor énfasis en aspectos superiores de escritura; a saber: adecuación, desarrollo de ideas, coherencia

y cohesión, más tres relacionados con aspectos lingüísticos: léxico, gramática y ortografía (ver Anexo 2.6).

5.4.2.3. SEGUNDA PARTE: CORRECCIÓN CON LA SEGUNDA VERSIÓN DE LA RÚBRICA

Una vez afinada la rúbrica y corregidos los textos, se condujo el análisis estadístico. Antes de comenzar la descripción detallada de cada una de las tablas que FACETS arroja, es importante mencionar que el evaluador 4 fue eliminado de los siguientes análisis. Esto debido a que presentó una alta inconsistencia en sus evaluaciones, mismas que distorsionaban ampliamente el modelo de FACETS. Recordemos que Linacre (2012b, p. 11) menciona que en exámenes de alto impacto, los rangos aceptables se encuentran entre 0.8 y 1.2 para los índices de *Outfit* e *Infit*. En la Tabla 5.3 se aprecia que sus valores de *Infit* (1.53) y de *Outfit* (1.64) se encuentran por afuera de los rangos aceptables. También se puede apreciar que es la evaluadora más indulgente y con correlaciones por debajo de las esperadas (0.37 en vez de 0.63 –predichas por el modelo–). De igual forma, su porcentaje de observaciones exactas se encuentran debajo de las esperadas (38.1 en vez de 45.3). Linacre (2012c) sugiere que cuando estos casos se presenten, se elimine al evaluador que está distorsionando la evaluación y se corra nuevamente el análisis.

Tabla 5.3: La evaluadora 4 es asistemática (segunda versión de la rúbrica)

Total Score	Total Count	Obsvd Average	Fair(M) Average	Model Measure	S.E.	Infit MnSq	ZStd	Outfit MnSq	ZStd	Estim. Discrm	Correlation PtMea	PtExp	Exact Obs %	Agree. Exp %	N Reader
1634	728	2.24	2.31	-.61	.06	1.53	8.8	1.64	9.0	.32	.37	.63	38.1	45.3	4 4
1527	728	2.10	2.15	-.20	.06	.98	-.3	.99	-.1	1.04	.67	.64	47.3	46.0	5 5
2521	1246	2.02	2.05	.03	.05	.82	-5.1	.82	-4.9	1.23	.71	.63	49.6	45.5	2 2
1456	735	1.98	2.01	.11	.06	.92	-1.6	.90	-1.8	1.12	.71	.64	50.0	45.8	3 3
1408	728	1.93	1.95	.24	.06	.97	-.5	1.00	.0	.97	.59	.65	45.4	45.2	6 6
1981	1064	1.86	1.87	.43	.05	.95	-1.3	.94	-1.4	1.09	.69	.64	48.4	44.1	1 1
1754.5	871.5	2.02	2.06	.00	.06	1.03	.0	1.05	.1		.62				Mean (Count: 6)
390.4	207.3	.12	.14	.33	.01	.23	4.3	.27	4.3		.12				S.D. (Population)
427.6	227.0	.13	.16	.36	.01	.25	4.7	.30	4.7		.13				S.D. (Sample)

Model, Populn: RMSE .06 Adj (True) S.D. .33 Separation 5.82 Strata 8.09 Reliability (not inter-rater) .97
 Model, Sample: RMSE .06 Adj (True) S.D. .36 Separation 6.39 Strata 8.85 Reliability (not inter-rater) .98
 Model, Fixed (all same) chi-square: 198.9 d.f.: 5 significance (probability): .00
 Model, Random (normal) chi-square: 4.9 d.f.: 4 significance (probability): .30
 Inter-Rater agreement opportunities: 8890 Exact agreements: 4164 = 46.8% Expected: 4023.8 = 45.3%

Severo desajuste del evaluador 4

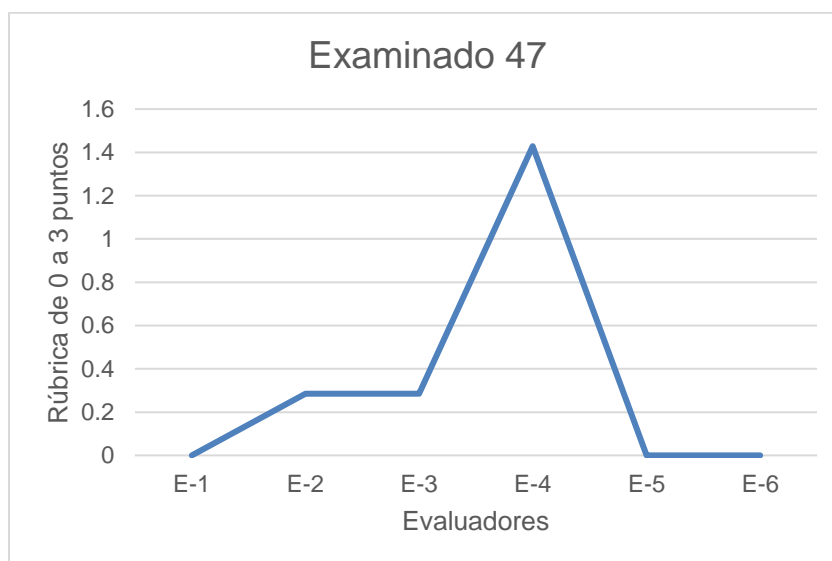
A continuación, se muestran un par de ejemplos que demuestran la falta de congruencia en las evaluaciones de la evaluadora 4. Como se puede apreciar en el Cuadro 5.12, la correctora fue sumamente indulgente con la sustentante. Mientras todos los evaluadores le asignaron cero a la mayoría de los criterios, ella asignó 1 e incluso 2 a tres de las categorías. En la Gráfica 5.1 se observa como la media del examinado 47 difiere mucho entre la evaluadora 4 y el resto. Los

evaluadores 1, 5 y 6 coinciden en que la competencia de este examinado es 0. Los evaluadores 2 y 3 le asignan un punto al léxico y a la ortografía (media de 0.28), mientras que la media del evaluador 4 es de 1.42, muy por encima de los demás evaluadores. Esto demuestra una severa inconsistencia en la evaluación.

Cuadro 5.12: La evaluadora 4 es sumamente indulgente

Examinado	Evaluador	Tarea	Adec	Ideas	Cohes	Coher	Léxico	Gram	Ortog
47	1	T2	0	0	0	0	0	0	0
	2	T2	0	0	0	0	1	0	1
	3	T2	0	0	0	0	1	0	1
	4	T2	1	1	2	1	1	2	2
	5	T2	0	0	0	0	0	0	0
	6	T2	0	0	0	0	0	0	0

Gráfica 5.1: Medias de los evaluadores para el examinado 47



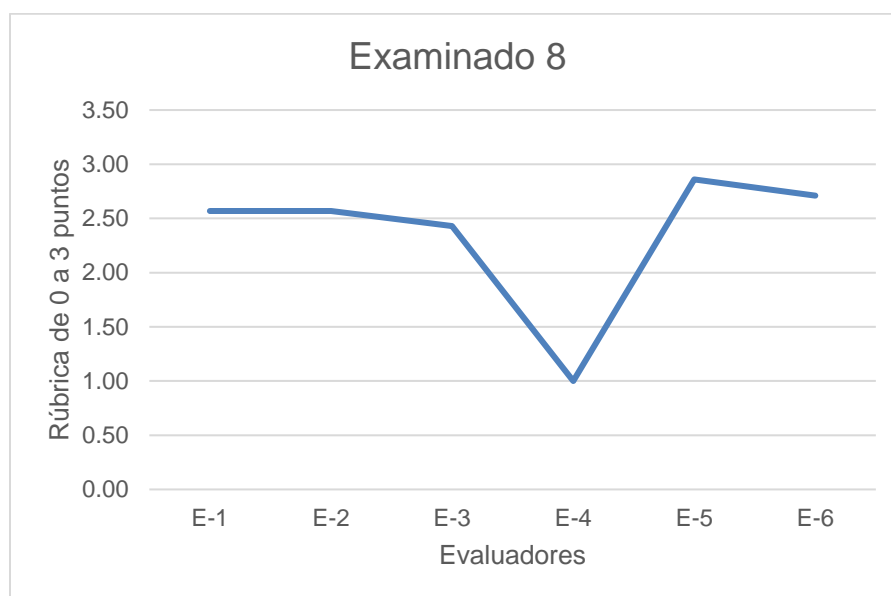
Sin embargo, el mayor problema estriba en la falta de sistematización de la evaluadora 4. Como podemos ver en el Cuadro 5.13, la evaluadora es sumamente severa con el examinado 8. Destaca claramente que el nivel del candidato se encuentra entre las bandas 2 y 3 en las distintas categorías, no obstante, la evaluadora 4 lo ubica en el nivel 1 en todas las categorías. Aunado a la severidad y falta de congruencia en el uso de la rúbrica, se evidencia un marcado efecto halo al asignar el nivel 1 a todas las categorías. Las medias de los demás evaluadores oscilaron entre 2.43 y 2.86. La Gráfica 5.2 muestra nuevamente una diferencia considerable entre las medias de

la evaluadora 4 y el resto; pero en esta ocasión, la evaluadora es muy severa y con un denotado efecto halo en su evaluación.

Cuadro 5.13: La evaluadora 4 es sumamente severa

Examinado	Juez	Tarea	Adec	Ideas 1	Ideas 2	Cohes	Coher	Léxico	Gram	Ortog
8	1	T1	3	2		3	2	2	3	3
	2	T1	2	2		3	3	2	2	3
	2	T2	2		2	3	3	3	3	3
	3	T1	2	2		3	2	2	3	3
	4	T2	1		1	1	1	1	1	1
	5	T2	3		3	3	3	3	2	3
	6	T1	3	3		3	3	2	2	3

Gráfica 5.2: Medias de los evaluadores para el examinado 8



Más aún, si contrastamos los resultados de la evaluadora 4 para los alumnos 47 y 8, observamos que ésta le otorgó un resultado más alto al alumno 47 (media 1.43) en comparación con el 8 (media 1). Sin embargo, al comparar los textos de ambos candidatos, podemos ver claramente la falta de consistencia y de sistematización en la asignación de los niveles de las distintas categorías.

Texto del candidato 47

La importancia del tema es que la diferente entre la educacion de las mujeres y de los hombres. Las mujeres beneficios mujer educacion que las hombres, des pues la primera. Tambien, los hombres no tienen no muchos sin estudios comparte a las mujeres. Es posible que los hombres trabaja antes las

mujeres. Creo que ahora, es diferente. Pero, remar que que de pues la clase preparatoria, muchos estudiante no continuar.

Texto del candidato 8

La idea que “la educación por sí misma amplía los horizontes de los individuos y constituye un beneficio para éstos” tiene un fundamento importante para nuestras sociedades y forma parte de nuestras culturas. La posibilidad de estudiar abre las puertas para las personas que no quieren estar contentos con su situación de vida, sino quieren aspirar a mejorar sus condiciones de vida, sus perspectivas y asimismo las perspectivas de sus hijos. Por otro lado, la idea de Milosavljevic de que “la situación educativa de hombres y mujeres tiene diferentes beneficios para unos y otros y para sus hijos e hijas” sigue siendo relevante en la actualidad.

Originalmente, el privilegio de estudiar se les possibilitó a los hombres sin que las mujeres tuvieran acceso para realizar sus propios intereses y aspiraciones. El papel de las mujeres se reducía a quedarse en casa, cuidar a los niños y preparar la comida. Cuando surgieron las oleadas del feminismo, uno de los factores más importantes aparte de un sufragio femenino universal y salario igualitario que requerían las mujeres era asegurar y garantizar la educación para las mujeres.

Aunque se ha logrado bastante acerca de mejorar el estatus de las mujeres dentro del sistema educativo, todavía sobran discrepancias en acceso a la educación entre hombres y mujeres sobre todo en algunos país de Tercer Mundo con una tradición machista.

Mientras en Europa o Estados Unidos, incluso a la República Checa, ocurre el problema contrario cuando los hombres no tienen tantas capacidades para estudiar como las mujeres. Por lo tanto, es obvio que las mujeres estudiantes pueden fácilmente superar en números a los hombres estudiantes. No obstante, cuando las mujeres luego entran al trabajo, siguen siendo discriminadas acerca del salario, promoción, liderazgo etc.

Sin embargo, estas tendencias cuando las mujeres superan en cifras a los hombres se refiere también a los países latinoamericanos por ejemplo. Mientras las mujeres están ingresadas en las universidades, los hombres famosos por ser Ni-Ni (ni estudiando, ni trabajando) desperdician sus valores en el mercado laboral.

Desde mi punto de vista y mi experiencia de la República Checa diría que tendríamos que asegurar que las mujeres estudian y luego trabajan en las mismas condiciones como los hombres y que comparten sus obligaciones de la manera igualitaria aunque por cierto existen otras diferencias biológicas¹⁸.

Cabe mencionar que cuando se comenzaron a identificar este tipo de inconsistencias con el primer paquete que se le otorgó a la evaluadora, se le pidió que corrigiera nuevamente algunos textos considerando que había sido demasiado severa o indulgente. No obstante, los resultados que se tomaron en cuenta para el estudio provinieron del primer resultado otorgado por el

¹⁸ Este texto originalmente fue escrito a mano; sin embargo, para su mejor comprensión se transcribió tal cual lo redactó el sustentante.

evaluador. Debido a que no se presentaban mejoras en la calidad de la evaluación, se la convocó para una sesión de revisión y de retroalimentación con el investigador. En primer lugar, se le pidió que evaluara nuevamente el texto del examinado 8 y que verbalizara por qué asignaba los valores que asignaba, de acuerdo con los criterios descritos en cada una de las categorías de la rúbrica. Nuevamente, los puntajes otorgados al texto del examinado 8 fueron muy bajos. Curiosamente, cuando se le pidió que argumentara la razón de éstos, no había ninguna explicación clara. De hecho, mayoritariamente la respuesta era: “no me convence”. Por ejemplo, en la ortografía se puede apreciar que el texto carece 100% de faltas de ortografía, sin embargo, la evaluadora le otorgó un 1 a esta categoría. Cuando se le preguntó que verbalizara los errores ortográficos del candidato, nuevamente la respuesta fue: “no tiene, pero no me convence”. Durante esta sesión de dos horas, la falta de entendimiento de los criterios de evaluación se hizo claramente evidente. Pese a la retroalimentación, sus evaluaciones subsecuentes continuaron siendo inestables y poco confiables, como se pudo ver en los análisis estadísticos.

Una vez que se eliminaron los datos de la evaluadora 4, se corrió nuevamente el análisis estadístico en FACETS. En la Figura 5.6 se observa el mapa de variabilidad. Lo primero que podemos observar es que al modificar los valores de la escala en un rango de 0 a 3, la escala en lógitos también aumentó en un rango de -5 a +5 lógitos, lo cual significa que nos permite identificar con mayor precisión la habilidad de los examinados. En la segunda columna se observan los 100 examinados, representados cada uno por un asterisco. Nuevamente, podemos observar una tendencia por arriba de la media (0 en lógitos), lo cual indica que la habilidad de escritura de la mayoría de los examinados es superior a la media. Esto tiene sentido, pues al igual que en la primera evaluación (pilotaje) la muestra se conformó por estudiantes de grado y posgrado matriculados en alguna universidad. Las siguientes facetas: evaluador, tarea y categorías de la rúbrica se encuentran invertidas, por lo cual, los valores por arriba del 0 indican mayor severidad, mayor dificultad de la tarea y también de la categoría de la rúbrica. Sin embargo, se puede apreciar que, a diferencia de los resultados del pilotaje, los datos se encuentran concentrados en el centro, lo cual indica rangos aceptables de severidad de los evaluadores o dificultad de la tarea/categorías de la rúbrica. A excepción de la categoría 7 (ortografía), los demás valores de encuentran próximos al centro de la escala. En la última columna se aprecia el cambio de niveles de la rúbrica de 0 a 3 puntos. Nuevamente, la línea punteada entre estos rangos denota el nivel umbral en el cual un candidato “x” habría tenido la misma posibilidad de haber caído en un nivel o en otro.

Figura 5.6: Mapa de variabilidad de corrección con la segunda versión de la rúbrica

Measr	+examinee	-Reader	-Essay	-Rubric categories	Scale
5	+	+	+	+	(3)
	*				
	**				
4	+	+	+	+	
	**				
	*				
	**				

3	+	+	+	+	
	**				
	***				---

2	+	+	+	+	
	**				

	*****				2
1	+	+	+	+	

	*****	1		4	
	*****	6	1	3 5 6	
*	*	* 2 3	*	* 1 2	* --- *
	****		2		
	***	5			

	*				
-1	+	+	+	7	
	*				1
	*				

-2	+	+	+	+	
	*				

-3	+	+	+	+	
-4	+	+	+	+	
-5	+	+	+	+	(0)

Nomenclatura

Primera columna
Measr= medición, dada en lógitos.

Segunda columna
Examinee= examinados
* indica un candidato

Tercera columna
Reader= evaluador

Cuarta columna
Essay= texto
1 – descripción de gráfica
2 – redacción de ensayo argumentativo

Quinta columna
Rubric categories = categorías de la rúbrica
1 – Adecuación
2 – Desarrollo de ideas
3 – Cohesión
4 – Coherencia
5 – Léxico
6 – Gramática
7 – Ortografía

Sexta columna
Niveles de la rúbrica (0 a 3 puntos)

El mapa de variabilidad nos ayuda a tener una primera impresión de la severidad o dificultad de las facetas; sin embargo, no nos proporciona detalles sobre la calidad en cada una de éstas, para ello, se requieren las tablas con los reportes de medición de la calidad de cada una de las facetas. Los derivados de las Tablas 5.4 muestran esta información; a saber: los examinados (Tabla 5.4a), los evaluadores (Tabla 5.4b), las tareas (Tabla 5.4c) y las categorías de la rúbrica (Tabla 5.4d).

En cuanto a la Tabla 5.4a (examinados) uno de los datos más importantes que arroja este reporte de medición es el puntaje justo que merece el examinado. En esta tabla podemos ver los datos jerarquizados de mayor a menor habilidad. Este dato aparece en dos formas: en lógitos y también en el promedio de los valores observados. Por ejemplo, el primer dato que figura en la Tabla 5.4a es el del candidato con menores habilidades de escritura (examinado 47). El candidato fue observado 35 veces (7 categorías de la rúbrica y 5 evaluadores). En todas estas observaciones, solo recibió 4 puntos, por lo que su habilidad es de 0.11 (en la escala del 0 al 3) o de - 4.92 lógitos (en la escala que FACETS establece). Debido a que la escala en lógitos suele ser de poca utilidad para el evaluador, el programa ofrece el puntaje con base en la escala de la rúbrica. Sin embargo, también ofrece el puntaje justo. En este caso, el programa determina que el puntaje justo para este examinado es de 0.09 incluso menor a lo observado. En el caso opuesto, al final de la Tabla 5.4a, se ubica el candidato con mayor habilidad (examinado 9). Este sustentante fue observado 63 veces (7 categorías de la rúbrica y 5 evaluadores –aunque cuatro de ellos tanto en la tarea 1 como en la 2–). Su puntaje observado es de 2.86 y el que sugiere el modelo es 2.87 (en la escala del 0 al 3) o de 4.36 lógitos.

Otro dato que resulta relevante en este reporte de los examinados es el valor de *Outfit* e *Infit*. En este caso, estos valores nos proporcionan indicios de variación en los resultados observados en el candidato. Recordemos que Linacre (2012b) menciona que los valores menores a 0.8 o mayores a 1.2 indican una inconsistencia. Aquellos mayores a 1.5 dan indicio de una severa inconsistencia. En el caso de los examinados, significaría que sus habilidades no son homogéneas o que presentan perfiles desiguales de escritura. En el caso de que los resultados del examen sirvieran como diagnóstico para cursos, el resultado de estos estudiantes sería de gran utilidad para el diseño de cursos enfocados al desarrollo de determinadas habilidades de escritura.

Tabla 5.4a: Reporte de medición de los candidatos con la segunda versión de la rúbrica

Total Score	Total Count	Obsvd Average	Fair (M) Average	Model		Infit		Outfit		Estim. Discrm	Correlation		Num examinee
				Measure	S.E.	MnSq	ZStd	MnSq	ZStd		PtMea	PtExp	
4	35	.11	.09	-4.92	.53	.91	.0	.83	-.1	1.06	.33	.20	47 47
26	49	.53	.51	-2.69	.26	.83	-.9	.84	-.9	1.26	.53	.31	61 61
58	63	.92	.90	-1.54	.21	1.42	2.1	1.43	2.2	.53	-.08	.30	12 12
34	35	.97	.90	-1.53	.28	1.23	.9	1.22	.9	.79	.29	.30	63 63
44	49	.90	.90	-1.53	.24	1.64	2.8	1.63	2.8	.28	.36	.32	49 49
45	49	.92	.92	-1.48	.24	1.02	.1	1.01	.1	.96	-.03	.32	50 50
51	49	1.04	1.04	-1.14	.24	1.06	.3	1.05	.3	.95	.60	.33	29 29
48	42	1.14	1.13	-.88	.25	.73	-1.2	.73	-1.2	1.28	.55	.29	91 91
60	49	1.22	1.22	-.65	.23	.65	-1.8	.64	-1.9	1.37	.54	.34	35 35
52	42	1.24	1.23	-.63	.25	.79	-.9	.81	-.8	1.22	.26	.30	92 92
55	42	1.31	1.28	-.51	.25	.77	-1.1	.77	-1.1	1.25	.43	.29	88 88
54	42	1.29	1.28	-.51	.25	1.15	.7	1.15	.7	.86	.32	.29	89 89
49	35	1.40	1.32	-.40	.27	.54	-2.3	.53	-2.3	1.51	.58	.31	45 45
58	42	1.38	1.35	-.33	.25	.87	-.5	.87	-.5	1.13	.33	.29	90 90
50	35	1.43	1.35	-.33	.27	1.04	.2	1.04	.2	.96	.37	.31	43 43
67	49	1.37	1.37	-.28	.23	.78	-1.1	.79	-1.1	1.22	.51	.34	25 25
68	49	1.39	1.39	-.23	.23	1.35	1.6	1.36	1.6	.62	.47	.34	23 23
52	35	1.49	1.41	-.18	.27	1.44	1.7	1.44	1.7	.54	.01	.31	54 54
68	49	1.39	1.43	-.11	.23	1.85	3.5	1.85	3.5	.09	.34	.31	99 99
76	49	1.55	1.53	.13	.23	1.23	1.1	1.24	1.2	.74	.03	.33	60 60
57	35	1.63	1.55	.17	.27	1.63	2.4	1.63	2.4	.26	-.12	.31	40 40
57	35	1.63	1.55	.17	.27	1.18	.8	1.19	.8	.81	.10	.31	52 52
57	35	1.63	1.55	.17	.27	1.25	1.1	1.26	1.1	.71	-.01	.31	55 55
77	49	1.57	1.57	.24	.23	.76	-1.3	.76	-1.3	1.29	.54	.33	41 41
58	35	1.66	1.58	.25	.27	1.11	.5	1.11	.5	.89	.18	.31	58 58
81	49	1.65	1.63	.39	.23	1.21	1.0	1.21	1.1	.77	-.17	.33	62 62
115	70	1.64	1.64	.41	.19	.97	-.1	.97	-.1	1.04	.11	.32	14 14
55	35	1.57	1.65	.43	.27	.71	-1.3	.70	-1.3	1.34	.48	.31	95 95
81	49	1.65	1.66	.44	.23	.79	-1.1	.79	-1.1	1.26	.45	.31	26 26
81	49	1.65	1.66	.44	.23	.83	-.8	.83	-.8	1.20	.23	.31	30 30
82	49	1.67	1.68	.49	.23	1.56	2.5	1.56	2.5	.34	.62	.34	27 27
62	35	1.77	1.69	.53	.27	.88	-.5	.88	-.4	1.14	.15	.30	64 64
84	49	1.71	1.74	.65	.23	1.34	1.6	1.34	1.6	.60	.14	.34	87 87
64	35	1.83	1.75	.67	.27	1.09	.4	1.10	.5	.89	.11	.30	59 59
59	35	1.69	1.77	.71	.27	1.04	.2	1.05	.2	.96	.60	.31	96 96
87	49	1.78	1.78	.75	.23	.84	-.8	.84	-.8	1.20	.57	.34	39 39
86	49	1.76	1.78	.75	.23	.70	-1.7	.71	-1.6	1.34	.12	.31	73 73
89	49	1.82	1.82	.85	.23	.83	-.8	.85	-.7	1.19	.26	.33	42 42
89	49	1.82	1.82	.85	.23	.94	-.2	.94	-.2	1.09	.60	.34	31 31
79	42	1.88	1.88	.99	.24	.82	-.8	.83	-.8	1.20	.24	.32	21 21
93	49	1.90	1.88	1.00	.23	.95	-.1	.96	-.1	1.06	.18	.33	56 56
92	49	1.88	1.89	1.01	.23	.85	-.7	.85	-.7	1.18	.38	.29	100 100
91	49	1.86	1.89	1.01	.23	.91	-.4	.91	-.4	1.11	.33	.31	98 98
92	49	1.88	1.91	1.06	.23	1.74	3.2	1.74	3.2	.11	.08	.30	65 65
94	49	1.92	1.93	1.10	.23	.69	-1.7	.69	-1.7	1.36	.22	.31	24 24
65	35	1.86	1.94	1.14	.27	.55	-2.3	.55	-2.3	1.51	.29	.30	94 94
83	42	1.98	1.97	1.23	.25	.74	-1.3	.74	-1.3	1.28	.24	.28	13 13
97	49	1.98	1.99	1.26	.23	1.06	.3	1.06	.3	.92	.05	.30	38 38
84	42	2.00	2.00	1.29	.25	1.48	2.0	1.50	2.1	.42	-.09	.28	16 16
98	49	2.00	2.01	1.31	.23	1.11	.6	1.10	.5	.88	.34	.30	34 34
85	42	2.02	2.02	1.35	.25	1.10	.5	1.09	.5	.88	.38	.32	19 19
84	42	2.00	2.03	1.36	.25	.79	-1.0	.80	-.9	1.26	.44	.34	6 6
98	49	2.00	2.03	1.37	.23	1.00	.0	.99	.0	1.00	.22	.30	71 71
74	35	2.11	2.04	1.40	.27	1.25	1.1	1.24	1.0	.69	.22	.29	51 51

74	35	2.11	2.04	1.40	.27	1.25	1.1	1.24	1.0	.69	.22	.29	51	51
100	49	2.04	2.07	1.48	.23	.87	-.6	.87	-.6	1.18	.55	.33	77	77
103	49	2.10	2.13	1.64	.23	1.02	.1	1.02	.1	.99	.40	.32	79	79
103	49	2.10	2.13	1.64	.23	.69	-1.8	.69	-1.8	1.43	.74	.32	85	85
104	49	2.12	2.15	1.69	.23	.75	-1.3	.75	-1.4	1.30	.30	.29	75	75
78	35	2.23	2.16	1.71	.28	.77	-1.0	.79	-.9	1.24	-.11	.28	48	48
73	35	2.09	2.17	1.72	.27	.74	-1.2	.74	-1.1	1.34	.52	.29	93	93
93	42	2.21	2.22	1.85	.25	.96	-.1	.96	-.1	1.00	-.09	.27	18	18
94	42	2.24	2.24	1.92	.26	.82	-.9	.85	-.7	1.21	.27	.30	17	17
81	35	2.31	2.25	1.94	.28	.55	-2.4	.56	-2.4	1.60	.33	.27	44	44
110	49	2.24	2.26	1.96	.24	1.65	3.0	1.62	2.9	.18	.24	.29	68	68
111	49	2.27	2.28	2.01	.24	.74	-1.5	.74	-1.5	1.35	.41	.29	28	28
113	49	2.31	2.32	2.13	.24	1.04	.2	1.04	.2	.94	.21	.31	76	76
97	42	2.31	2.34	2.18	.26	1.12	.6	1.11	.5	.81	.34	.28	3	3
114	49	2.33	2.34	2.19	.24	.90	-.5	.92	-.3	1.12	.27	.28	66	66
84	35	2.40	2.34	2.19	.29	.75	-1.2	.79	-.9	1.28	.07	.26	46	46
114	49	2.33	2.36	2.24	.24	.80	-1.1	.81	-1.0	1.26	.27	.27	22	22
115	49	2.35	2.36	2.25	.24	.82	-1.0	.84	-.8	1.25	.43	.28	36	36
116	49	2.37	2.38	2.31	.25	.80	-1.1	.78	-1.2	1.32	.53	.30	78	78
99	42	2.36	2.39	2.33	.26	.84	-.7	.81	-.9	1.30	.67	.31	4	4
117	49	2.39	2.40	2.36	.25	1.20	1.1	1.18	1.0	.73	.27	.27	32	32
118	49	2.41	2.42	2.43	.25	1.51	2.5	1.47	2.3	.31	.21	.27	72	72
102	42	2.43	2.43	2.46	.27	.81	-1.0	.81	-.9	1.27	.30	.24	20	20
153	63	2.43	2.44	2.47	.22	.91	-.5	.88	-.7	1.15	.37	.28	10	10
118	49	2.41	2.44	2.48	.25	.82	-1.0	.81	-1.0	1.27	.40	.26	69	69
119	49	2.43	2.45	2.50	.25	.84	-.9	.81	-1.0	1.26	.43	.29	33	33
120	49	2.45	2.47	2.56	.25	.98	.0	1.01	.1	1.00	.22	.29	84	84
122	49	2.49	2.51	2.68	.26	1.81	3.7	1.79	3.6	-.06	.15	.26	74	74
104	42	2.48	2.51	2.68	.28	.89	-.5	1.00	.0	1.02	-.09	.26	7	7
124	49	2.53	2.55	2.82	.26	.95	-.2	.91	-.4	1.07	.23	.27	83	83
87	35	2.49	2.56	2.85	.30	.81	-.8	.87	-.5	1.19	.10	.25	97	97
107	42	2.55	2.58	2.92	.29	.98	.0	1.09	.5	.89	-.14	.25	1	1
107	42	2.55	2.58	2.92	.29	.69	-1.7	.67	-1.7	1.43	.47	.25	5	5
108	42	2.57	2.58	2.94	.29	1.11	.6	1.05	.3	.90	.29	.25	15	15
127	49	2.59	2.61	3.03	.27	.87	-.7	.88	-.5	1.13	.17	.24	70	70
93	35	2.66	2.62	3.07	.34	1.03	.2	1.18	.7	.85	-.15	.21	57	57
128	49	2.61	2.63	3.12	.28	.77	-1.2	.69	-1.6	1.35	.61	.25	37	37
110	42	2.62	2.65	3.19	.30	.81	-.9	.77	-1.0	1.23	.34	.26	8	8
111	42	2.64	2.67	3.27	.31	1.26	1.2	1.17	.8	.77	.19	.22	11	11
112	42	2.67	2.70	3.38	.32	.79	-.9	.73	-1.1	1.24	.41	.24	2	2
132	49	2.69	2.71	3.45	.30	.83	-.7	.80	-.8	1.17	.34	.23	86	86
133	49	2.71	2.75	3.60	.31	.76	-1.1	.66	-1.4	1.27	.50	.22	82	82
135	49	2.76	2.77	3.74	.33	.97	.0	1.10	.4	.97	.09	.21	80	80
136	49	2.78	2.80	3.89	.34	1.01	.1	1.32	1.1	.90	-.08	.18	67	67
100	35	2.86	2.84	4.16	.48	.97	.0	1.07	.2	.99	.08	.14	53	53
139	49	2.84	2.86	4.29	.38	1.00	.0	1.49	1.3	.93	-.02	.17	81	81
180	63	2.86	2.87	4.36	.36	.94	-.1	1.03	.1	1.02	.16	.16	9	9
88.9	45.0	1.97	1.97	1.28	.26	.99	-.1	1.00	-.1		.27		Mean (Count: 100)	
29.1	6.9	.55	.56	1.58	.05	.28	1.3	.29	1.3		.21		S.D. (Population)	
29.3	7.0	.55	.57	1.59	.05	.28	1.3	.29	1.4		.21		S.D. (Sample)	

Model, Populn: RMSE .27 Adj (True) S.D. 1.56 Separation 5.86 Strata 8.15 Reliability .97														
Model, Sample: RMSE .27 Adj (True) S.D. 1.56 Separation 5.89 Strata 8.19 Reliability .97														
Model, Fixed (all same) chi-square: 3004.8 d.f.: 99 significance (probability): .00														
Model, Random (normal) chi-square: 95.0 d.f.: 98 significance (probability): .57														

En este caso, vamos a observar el valor de *Outfit* más elevado. Éste le corresponde al examinado 99, con un valor de 1.85. Este dato nos da indicios de que los patrones del examinado son inestables. Veamos ahora el resultado que los evaluadores le otorgaron a este candidato. Como se puede apreciar en el Cuadro 5.14, existen variaciones considerables tanto en cada una de las categorías de la rúbrica como en los resultados otorgados por cada uno de los evaluadores. Las primeras dos categorías (registro académico y desarrollo de ideas) presentan valores de 0 hasta 3 puntos. La cohesión y la coherencia, rangos de 0 a 2 puntos, el vocabulario y la gramática 1 y 2 puntos, mientras que la ortografía 2 y 3 puntos. En primer lugar, estos resultados denotan un perfil desigual del sustentante (valores más altos en órdenes inferiores de escritura: vocabulario, gramática y ortografía, mientras que en órdenes superiores de escritura:

cohesión, coherencia, desarrollo de ideas y adecuación, valores más bajos). Sin embargo, también el perfil de los evaluadores es desigual. El rango varía desde 0.86 hasta 2.43 (en escala el 0 al 3).

Cuadro 5.14: Examinado con mayor índice de desajuste

Candi- dato	Eva- luador	Tarea	Categorías de la rúbrica							Pro- medio
			Adec	Ideas	Cohes	Coher	Vocab	Gram	Ortog	
99	1	T1	1	0	2	2	1	2	3	1.57
	1	T2	1	0	1	0	1	1	2	0.86
	2	T1	0	1	0	0	1	1	2	0.71
	2	T2	0	1	1	1	2	1	2	1.14
	3	T1	1	1	0	0	1	1	2	0.86
	5	T1	3	3	2	2	2	2	3	2.43
	6	T1	3	2	2	2	2	2	2	2.14
Promedio por categoría			1.29	1.14	1.14	1.00	1.43	1.43	2.29	

Veamos ahora los textos producidos por el candidato 99.

Texto1: Descripción y contraste de gráficos

Esta gráfica del INEGI presenta el cambio del nivel de estudios de los mexicanos relacionando con la población entre 1999 y 2009. De promera vista, se ve que el nivel de escolarización bajó entre estos diez años, sin embargo, esta disminución es proporcionada dado que el promedio de hijos de mujeres desminuyó 0.5 por ciento del 1999 y 2009. El porcentaje de las personas sin ninguna escolaridad disminuyó en estos diez años y subió la cantidad de personas que llevan estudios superiores.

Texto 2: Ensayo argumentativo

El cuidado de salud es un asunto muy importante para bienestar de una persona de cualquier parte del mundo, por lo tanto un gobierno que cuida la salud de su población demuestra su nivel de desarrollo del país. Siempre hay que ofrecer apoyo a las personas que cuentan poco recurso económico. En Japón existe el sistema de salud tanto pública como privada. Las personas pueden escoger dependiendo de su necesidad y de su situación económica. A mí parecer un gran porcentaje de los gastos médicos hay que cubrir de los impuestos. Si el sistema de salud pública ofrece buen servicio, las personas no utilizarían los privados. Los sistemas privados convierten la salud en negocio.

Cuando se encuentran discrepancias tan altas entre los evaluadores y en el perfil de los sustentantes, sería muy recomendable analizar los textos con detenimiento en sesiones de estandarización y homogenización de criterios con los evaluadores. Por esta razón, FACETS representa una herramienta de gran utilidad para poder identificar aquellos textos que están resultando difíciles de evaluar. Vamos a ver ahora el contrario, el examinado con los valores de

Outfit más bajos. Éste es el candidato 5, con un valor de 0.67. El Cuadro 5.15 muestra los resultados de este candidato. Se observa una gran estabilidad tanto en el perfil del candidato como en el de los evaluadores. El rango de las categorías es de 2 a 3 puntos. Las medias de las categorías van de 2.17 (gramática) a 3.0 (ortografía), mientras que la de los evaluadores oscila entre 2.14 (evaluador 1) y 2.86 (evaluador 5). Este tipo de textos podrían emplearse en las guías o manuales de los evaluadores como ejemplos prototípicos de un cierto nivel de desempeño.

Cuadro 5.15: Examinado con menor índice de desajuste

Candi- dato	Eva- luador	Tarea	Categorías de la rúbrica							Pro- medio
			Adec	Ideas	Cohes	Coher	Vocab	Gram	Ortog	
5	1	T1	2	2	2	2	2	2	3	2.14
	2	T1	2	3	2	3	3	2	3	2.57
	2	T2	2	3	3	3	2	2	3	2.57
	3	T1	2	3	3	2	3	3	3	2.71
	5	T1	3	3	3	3	3	2	3	2.86
	6	T2	3	2	2	2	3	2	3	2.43
Promedio por categoría			2.33	2.67	2.50	2.50	2.67	2.17	3.00	

En cuanto a la Tabla 5.4b (evaluadores), tras haber eliminado al evaluador 4 por haber presentado severos índices de desajuste, se observan parámetros estables de ajuste. Excepto por el evaluador 6, quien presenta un mínimo desajuste en sus mediciones (*Outfit* > 1.2). En primer lugar, se aprecia que los evaluadores se encuentran jerarquizados del más indulgente al más severo. En este caso, el evaluador más laxo es el 5 (-0.37 lógitos) y el más severo es el 1 (0.34 lógitos). Estos valores nos proporcionan un rango de severidad bastante aceptable (0.71)¹⁹.

En la primera columna se aprecia el valor total de las asignaciones del evaluador (*total score*) y también el número de observaciones (*total count*). Por ejemplo, el evaluador 5 llevó a cabo 728 observaciones (7 categorías = 104 textos evaluados). El total de los valores asignados fue de 1527, lo cual da un promedio de 2.10 (de 0 a 3 puntos). El modelo ofrece el puntaje justo, que debería ser 2.14 (su media de evaluaciones). En la segunda columna se aprecia el índice de severidad en lógitos y también el error estándar. El error estándar de los evaluadores varía entre 0.05 y 0.06, lo cual da indicios de un alto grado de precisión en el modelo. Los índices de calidad *Outfit* e *Infit* se encuentran dentro de los parámetros aceptables (excepto el evaluador 6, quien

¹⁹ De acuerdo al modelo actual, el rango permisible sería la cuarta parte de la suma -en lógitos- del examinado más bajo (-4.92) y el más alto (4.36) = 2.32.

se encuentra ligeramente fuera de los rangos). En la quinta columna se reportan las correlaciones observadas o puntos biserials (*PtMea*) y las esperadas (*PtExp*). Nuevamente, se observa un alto grado de consistencia en las correlaciones de los evaluadores, excepto en el evaluador 6, quien se encuentra ligeramente por debajo de los valores observados. Recordemos que el modelo parte del supuesto que los evaluadores son “evaluadores independientes” y no máquinas, razón por la cual no se esperan correlaciones perfectas, sino más bien consistencia en aquellas predichas por el modelo. En la penúltima columna figuran las observaciones exactas (*Exact obs*) en contraste con las esperadas (*Agree exp*). Nuevamente, todos los evaluadores se encuentran por encima de los valores observados, excepto el evaluador 6, quien se encuentra ligeramente por debajo del porcentaje esperado. La última columna hace mención al evaluador (*reader*).

Tabla 5.4b: Reporte de medición de los evaluadores con la segunda versión de la rúbrica

Total Score	Total Count	Obsvd Average	Fair(M) Average	Model Measure	S.E.	Infit MnSq	ZStd	Outfit MnSq	ZStd	Estim. Discrm	Correlation PtMea	PtExp	Exact Obs %	Agree. Exp %	N Reader
1527	728	2.10	2.14	-.37	.06	1.02	.4	1.03	.6	.98	.70	.69	50.0	48.1	5 5
2521	1246	2.02	2.03	-.08	.05	.88	-3.2	.90	-2.6	1.14	.73	.69	52.0	48.0	2 2
1456	735	1.98	2.01	-.02	.06	.91	-1.7	.90	-1.9	1.13	.75	.69	51.6	48.2	3 3
1408	728	1.93	1.94	.13	.06	1.21	3.9	1.25	4.3	.69	.56	.70	45.3	47.8	6 6
1981	1064	1.86	1.86	.34	.05	1.03	.7	1.02	.4	.98	.70	.69	51.2	46.9	1 1
1778.6	900.2	1.98	2.00	.00	.06	1.01	.0	1.02	.2		.68				Mean (Count: 5)
423.5	215.9	.08	.09	.24	.01	.12	2.4	.13	2.5		.07				S.D. (Population)
473.5	241.4	.09	.10	.26	.01	.13	2.7	.14	2.8		.07				S.D. (Sample)

Model, Populn: RMSE .06 Adj (True) S.D. .23 Separation 3.93 Strata 5.57 Reliability (not inter-rater) .94
 Model, Sample: RMSE .06 Adj (True) S.D. .26 Separation 4.42 Strata 6.23 Reliability (not inter-rater) .95
 Model, Fixed (all same) chi-square: 84.7 d.f.: 4 significance (probability): .00
 Model, Random (normal) chi-square: 3.8 d.f.: 3 significance (probability): .28
 Inter-Rater agreement opportunities: 6363 Exact agreements: 3200 = 50.3% Expected: 3038.8 = 47.8%

ligero desajuste del evaluador 6

En la Tabla 5.4c se reporta la medición de las tareas que se emplearon. Tarea 1: descripción y contraste de gráficas, tarea 2: ensayo argumentativo. Recordemos que debido a que en la primera evaluación (pilotaje) se observaron índices de dificultad muy próximos entre ambas tareas, se decidió lograr una mayor interacción entre los evaluadores y se evitó que los evaluadores corrigieran los dos textos de un mismo candidato. En el caso de los evaluadores 1 y 2, quienes evaluaron en su gran mayoría ambos textos, se decidió hacerlo de forma aleatoria y con códigos, de tal suerte que el evaluador no reconociera el nombre del examinado en ambas tareas. En la segunda columna, se puede apreciar que la diferencia entre la tarea 1 y 2 es de 0.4 lógitos. Si bien este dato es ligeramente mayor que en la primera evaluación, se encuentra dentro de los parámetros aceptables. Consecuentemente, pese a que la tarea 1 (descripción de

gráficas) se la evaluó con mayor severidad con respecto al ensayo argumentativo, la diferencia es mínima (0.15 puntos en la escala de la rúbrica con rangos de 0 a 3 puntos).

Tabla 5.4c: Reporte de medición de las tareas con la segunda versión de la rúbrica

Total Score	Total Count	Obsvd Average	Fair (M) Average	Model Measure	S.E.	Infit MnSq	ZStd	Outfit MnSq	ZStd	Estim. Discrm	Correlation PtMea	PtExp	N Essay
4926	2499	1.97	2.07	-.20	.03	1.00	.0	1.01	.3	1.00	.70	.71	2 2
3967	2002	1.98	1.92	.20	.04	1.00	.0	.99	-.1	1.00	.68	.67	1 1
4446.5	2250.5	1.98	2.00	.00	.04	1.00	.0	1.00	.1		.69		Mean (Count: 2)
479.5	248.5	.01	.08	.20	.00	.00	.0	.01	.3		.01		S.D. (Population)
678.1	351.4	.01	.11	.28	.00	.00	.0	.01	.4		.02		S.D. (Sample)

Model, Populn: RMSE .04 Adj (True) S.D. .19 Separation 5.42 Strata 7.56 Reliability .97
 Model, Sample: RMSE .04 Adj (True) S.D. .28 Separation 7.73 Strata 10.64 Reliability .98
 Model, Fixed (all same) chi-square: 60.8 d.f.: 1 significance (probability): .00

En la Tabla 5.4d se observa el reporte de medición de cada una de las categorías de la rúbrica. La jerarquización va de la que se evalúa de forma más indulgente (ortografía con -1.03 lógitos) a la más severa (coherencia con 0.36 lógitos). Entonces, el rango de severidad es de 1.39 lógitos, el cual es aceptable según el modelo. En la primera columna se aprecian las 643 observaciones para cada categoría y el promedio observado, así como el ajustado por el modelo. La diferencia entre la ortografía y la coherencia es de 0.53 puntos (en la escala de 0 a 3 de la rúbrica), aunque la diferencia entre la ortografía y la adecuación (segunda categoría evaluada de forma más indulgente) es de 0.38. Por esta razón, en el mapa de Wright se percibe un gran distanciamiento de la categoría 7 en relación con el resto de las categorías. El error estándar es 0.07 (ligeramente mayor a 0.05, pero aceptable) y los índices de calidad de la evaluación (*Outfit* e *Infit*) demuestran un ligero desajuste de la categoría 7: ortografía. Este desajuste también se puede apreciar en las correlaciones o puntos biserials (*PtMea*), pues en esta categoría se encuentran ligeramente por debajo de lo esperado (0.59 en vez de 0.67). Sin embargo, todas las demás categorías presentan un óptimo índice de calidad y también de correlación entre categorías. La categoría 5 (léxico) muestra un valor de *Outfit* ligeramente bajo (0.73 cuando el mínimo debería ser 0.8), aunque Linacre (2012b) asevera que esto es normal cuando una de las categorías se encuentra por arriba de los valores establecidos. No obstante, esto no afecta de ninguna manera la medición. En la última columna se observan las categorías de la rúbrica ordenadas de acuerdo a sus índices de dificultad.

Tabla 5.4d: Reporte de medición de las categorías de la segunda versión de la rúbrica

Total Score	Total Count	Obsvd Average	Fair(M) Average	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	N Rubric categories
1488	643	2.31	2.38	-1.03	.07	1.18	3.0	1.24	3.3	.78	.59	.67	7 7
1273	643	1.98	2.00	.00	.07	1.08	1.4	1.03	.6	.93	.69	.69	1 1
1266	643	1.97	1.98	.03	.07	1.18	3.1	1.13	2.3	.83	.68	.69	2 2
1240	643	1.93	1.94	.15	.07	.99	-.2	1.00	.0	1.00	.68	.69	3 3
1224	643	1.90	1.91	.22	.07	.71	-5.9	.73	-5.3	1.33	.74	.69	5 5
1211	643	1.88	1.89	.27	.07	.86	-2.7	.87	-2.4	1.16	.72	.69	6 6
1191	643	1.85	1.85	.36	.07	1.03	.5	1.02	.3	.96	.68	.69	4 4
1270.4	643.0	1.98	1.99	.00	.07	1.00	-.1	1.00	-.2		.68		Mean (Count: 7)
92.8	.0	.14	.17	.43	.00	.16	3.1	.15	2.7		.04		S.D. (Population)
100.2	.0	.16	.18	.47	.00	.17	3.3	.17	3.0		.05		S.D. (Sample)

Model, Populn: RMSE .07 Adj (True) S.D. .43 Separation 6.41 Strata 8.88 Reliability .98
 Model, Sample: RMSE .07 Adj (True) S.D. .46 Separation 6.93 Strata 9.58 Reliability .98
 Model, Fixed (all same) chi-square: 262.6 d.f.: 6 significance (probability): .00
 Model, Random (normal) chi-square: 5.9 d.f.: 5 significance (probability): .32

El Cuadro 5.16 muestra un ejemplo de un examinado con habilidades de escritura elementales, pero con un alto puntaje en la categoría 7 (ortografía); lo cual denota que la categoría no ayuda en realidad a discriminar entre los sustentantes con alto y bajo dominio de escritura, pues su valor es demasiado elevado en comparación con el de las demás categorías.

Cuadro 5.16: Examinado con altos puntajes en ortografía

Candi- dato	Eva- luador	Tarea	Categorías de la rúbrica							Pro- medio
			Adec	Ideas	Cohes	Coher	Vocab	Gram	Ortog	
49	1	T1	1	0	0	0	1	1	3	0.86
	1	T2	1	0	1	2	1	1	1	1.00
	2	T1	0	1	1	0	1	0	2	0.71
	2	T2	0	1	1	0	1	1	2	0.86
	3	T1	0	0	0	0	1	1	3	0.71
	5	T2	0	1	0	0	1	0	1	0.43
	6	T2	2	1	2	2	1	2	2	1.71
Promedio por categoría			0.57	0.57	0.71	0.57	1.00	0.86	2.00	

Veamos ahora los textos producidos por el candidato 49.

Texto1: Descripción y contraste de gráficos

A los padres 62.2 por ciento no fueron a la escuela pero a las hijas sin estudio son solamente 26.3 por ciento, aproximadamente 35 por ciento bajé que los padres. También la gente de las hijas que van a la licenciatura, universidad son 6.4 por ciento , es más aproximadamente 6 por ciento que los padres , pero todavía su por ciento es está un poco bajo.

Texto 2: Ensayo argumentativo

Según el cuadro de tarea 1 ,la situación educativa de los hombres y las mujeres son diferentes. Hay mucha gente que no fue a la escuela en los padres pero las hijas de sin estudio son muy poca. Es que la epoca de los padres, no era facil estudiar en la escuela y tal vez no era necesario ir a la escuela . Al contrario, ahora es normal que las hijas vayan a la escuela para su educación y por estas cosas las mujeres también pueden trabajar en la social. No deben que estar en la casa como ama de casa, las mujeres pueden trabajar igual que los hombres. En mi país, casi todos los padres de máximo estudio es bachillerato y de los hijas es licenciatura. Por eso no se dice que mi país es igual de méxico.

La última tabla que reporta FACETS es de suma relevancia, pues en ésta se observan los sesgos. A diferencia de los índices de ajuste, la tabla de sesgos indica que los datos se ajustan al modelo de manera útil (Eckes, 2012, p. 278). La Tabla 5.5 muestra la interacción entre el examinado y el evaluador.

Tabla 5.5: Reporte de sesgos entre evaluador y examinado con la segunda versión de la rúbrica

Observd Score	Expctd Score	Observd Count	Obs-Exp Average	Bias Size	Model S.E.	t	d.f.	Prob.	Infit Outfit		examinee			Reader			
									MnSq	MnSq	Sq	Num	exa	measr	N	R	measr
4	12.44	7	-1.21	-3.24	.67	-4.81	6	.0030	.7	.7	340	40	40	.17	5	5	-.37
1	10.81	7	-1.40	-4.60	1.08	-4.25	6	.0054	.7	.4	427	27	27	.49	6	6	.13
12	17.83	7	-.83	-2.26	.60	-3.81	6	.0089	1.3	1.3	374	74	74	2.68	5	5	-.37
8	14.21	7	-.89	-2.25	.62	-3.64	6	.0108	.1	.1	434	34	34	1.31	6	6	.13
12	17.31	7	-.76	-2.02	.60	-3.40	6	.0145	.5	.5	415	15	15	2.94	6	6	.13
12	17.29	7	-.76	-2.01	.60	-3.38	6	.0149	.4	.4	372	72	72	2.43	5	5	-.37
6	11.43	7	-.78	-2.04	.63	-3.21	6	.0184	.3	.3	354	54	54	-.18	5	5	-.37
13	20.56	14	-.54	-1.43	.45	-3.20	13	.0069	1.0	1.0	199	99	99	-.11	2	2	-.08
6	11.34	7	-.76	-2.00	.63	-3.16	6	.0196	.8	.8	260	60	60	.13	3	3	-.02
12	16.91	7	-.70	-1.85	.60	-3.10	6	.0211	.4	.4	432	32	32	2.36	6	6	.13
8	13.23	7	-.75	-1.89	.62	-3.06	6	.0221	.1	.1	419	19	19	1.35	6	6	.13
21	28.22	14	-.52	-1.29	.42	-3.05	13	.0093	.7	.6	138	38	38	1.26	2	2	-.08
16	19.61	7	-.52	-1.90	.63	-3.01	6	.0237	.7	.8	467	67	67	3.89	6	6	.13
0	8.83	7	-1.26	-4.90	1.64	-3.00	6	.0241	.1	.1	423	23	23	-.23	6	6	.13
6	11.02	7	-.72	-1.89	.63	-2.98	6	.0245	.3	.3	343	43	43	-.33	5	5	-.37
14	18.31	7	-.62	-1.79	.60	-2.97	6	.0250	.1	.1	470	70	70	3.03	6	6	.13
10	14.92	7	-.70	-1.77	.60	-2.94	6	.0259	1.2	1.2	3	3	3	2.18	1	1	.34
10	14.87	7	-.70	-1.75	.60	-2.91	6	.0270	.9	.9	251	51	51	1.40	3	3	-.02
7	11.81	7	-.69	-1.78	.63	-2.84	6	.0297	.1	.1	431	31	31	.85	6	6	.13
7	11.78	7	-.68	-1.76	.63	-2.82	6	.0304	.1	.1	430	30	30	.44	6	6	.13
9	13.77	7	-.68	-1.71	.61	-2.81	6	.0307	.7	.7	256	56	56	1.00	3	3	-.02
4	8.69	7	-.67	-1.89	.67	-2.81	6	.0309	.6	.6	425	25	25	-.28	6	6	.13
3	7.75	7	-.68	-2.02	.72	-2.79	6	.0316	.7	.7	349	49	49	-1.53	5	5	-.37
4	8.57	7	-.65	-1.85	.67	-2.74	6	.0337	1.6	1.6	89	89	89	-.51	1	1	.34
9	13.64	7	-.66	-1.66	.61	-2.73	6	.0341	.4	.4	424	24	24	1.10	6	6	.13
8	12.44	7	-.63	-1.61	.62	-2.61	6	.0402	1.1	1.1	355	55	55	.17	5	5	-.37
11	15.30	7	-.61	-1.55	.60	-2.61	6	.0403	.9	1.0	4	4	4	2.33	1	1	.34
2	6.46	7	-.64	-2.10	.82	-2.55	6	.0434	1.1	1.1	429	29	29	-1.14	6	6	.13
10	14.15	7	-.59	-1.48	.60	-2.46	6	.0489	.5	.5	416	16	16	1.29	6	6	.13

Datos centrales no incluidos por ser menos representativos de los sesgos

17	12.84	7	.59	1.59	.66	2.41	6	.0523	.5	.5	396	96	96	.71	5	5	-.37
18	13.66	7	.62	1.75	.71	2.48	6	.0478	1.0	1.2	398	98	98	1.01	5	5	-.37
13	8.79	7	.60	1.51	.60	2.53	6	.0445	.4	.4	329	29	29	-1.14	5	5	-.37
27	20.95	14	.43	1.08	.42	2.54	13	.0246	1.0	1.0	62	62	62	.39	1	1	.34
18	13.48	7	.65	1.82	.71	2.57	6	.0422	1.8	1.8	216	16	16	1.29	3	3	-.02
17	12.43	7	.65	1.73	.66	2.63	6	.0389	.5	.5	459	59	59	.67	6	6	.13
17	12.36	7	.66	1.76	.66	2.67	6	.0370	.6	.6	227	27	27	.49	3	3	-.02
20	14.50	7	.79	2.85	1.07	2.67	6	.0368	.9	.8	334	34	34	1.31	5	5	-.37
19	13.77	7	.75	2.28	.81	2.82	6	.0303	1.1	1.6	387	87	87	.65	5	5	-.37
11	6.46	7	.65	1.70	.60	2.85	6	.0292	1.3	1.3	463	63	63	-1.53	6	6	.13
16	11.03	7	.71	1.82	.63	2.88	6	.0279	1.7	2.1	452	52	52	.17	6	6	.13
16	11.03	7	.71	1.82	.63	2.88	6	.0279	.6	.7	455	55	55	.17	6	6	.13
13	8.13	7	.70	1.76	.60	2.94	6	.0258	.3	.3	491	91	91	-.88	6	6	.13
17	11.23	7	.82	2.16	.66	3.28	6	.0168	.5	.5	458	58	58	.25	6	6	.13
15	9.14	7	.84	2.11	.61	3.44	6	.0138	.4	.4	499	99	99	-.11	6	6	.13
12	6.46	7	.79	2.05	.60	3.45	6	.0136	.5	.5	449	49	49	-1.53	6	6	.13
16	10.03	7	.85	2.17	.63	3.45	6	.0136	1.5	1.5	454	54	54	-.18	6	6	.13
28	19.54	14	.60	1.52	.43	3.55	13	.0036	1.1	1.1	60	60	60	.13	1	1	.34
17	10.52	7	.93	2.41	.66	3.67	6	.0105	.5	.5	399	99	99	-.11	5	5	-.37
23	11.93	14	.79	2.08	.42	4.93	13	.0003	1.6	1.7	412	12	12	-1.54	6	6	.13
17.8	17.79	9.0	.00	.04	.65	-.02			.7	.7	Mean (Count: 500)						
8.8	8.21	3.2	.33	1.00	.24	1.46			.4	.5	S.D. (Population)						
8.8	8.22	3.2	.34	1.00	.24	1.46			.4	.5	S.D. (Sample)						

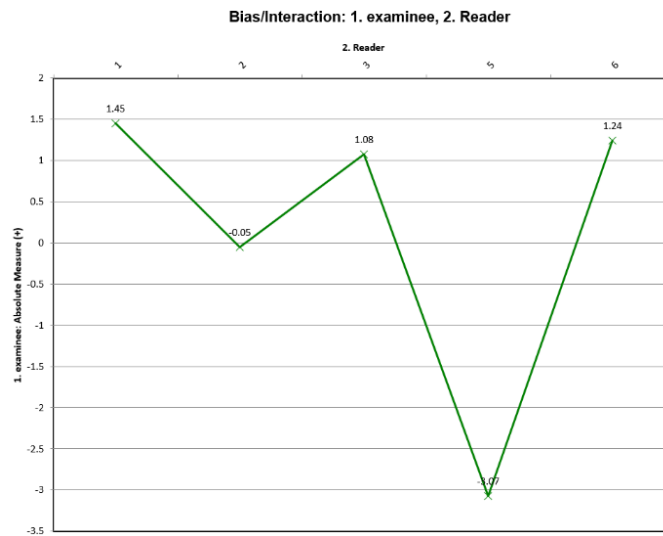
Fixed (all = 0) chi-square: 1069.4 d.f.: 500 significance (probability): .00

Los sesgos se encuentran jerarquizados por importancia; es decir, los que figuran en los extremos superior e inferior son los más significativos. En la primera columna encontramos cuatro elementos: el puntaje observado, el esperado, el total de observaciones y el promedio del puntaje observado menos el esperado. Si este último puntaje es positivo, entonces el evaluador es más indulgente de lo esperado. Si es negativo, significa que el evaluador fue más severo. En realidad, se espera que el último elemento de la tabla sea cero, pues los resultados observados deberían ser muy similares a los esperados. En la segunda columna encontramos el tamaño del sesgo (*bias size*), el error estándar del modelo (*model S.E.*), la *t* (*student's t-statistic testing the hypothesis*), los grados de libertad (*d.f.*) y la probabilidad. El tamaño del sesgo se encuentra en lógitos, mientras que el error estándar en un rango de 0 a 1. La *t* significa la hipótesis de que el sesgo sea una cuestión azarosa o no. Entre más se aleja del 0, más contundente es la probabilidad de que el sesgo no sea mera casualidad. Por esta razón, los datos se encuentran ordenados con respecto a la *t*. La tercera columna nos muestra los valores *Infit MnSq* y *Outfit MnSq*. En la última columna figuran datos sobre el examinado: el cuadrado, el examinado y su medición en lógitos, así como los datos relativos al examinador: el evaluador y su medición de severidad en lógitos.

Veamos ahora los dos ejemplos que figuran con los sesgos más altos (positivos y negativos). En la primera fila de la Tabla 5.5, observamos que el evaluador 5, quien además es

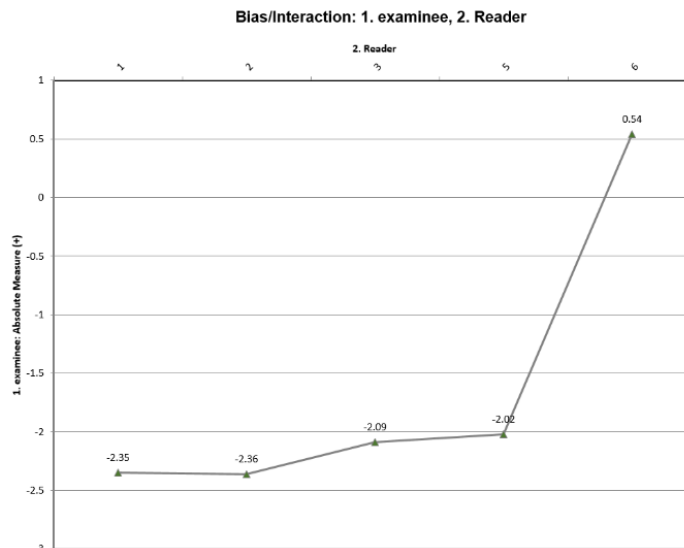
el que sistemáticamente es más indulgente (-0.37 lógitos) evaluó con gran dureza al examinado 40. Le otorgó solamente 4 puntos cuando el puntaje esperado era de 12.44 (predicho por el modelo). En la Gráfica 5.3 se aprecia claramente que el evaluador 5 fue excesivamente severo con el candidato 40.

Gráfica 5.3: Sesgos e interacciones entre evaluadores y el examinado 40



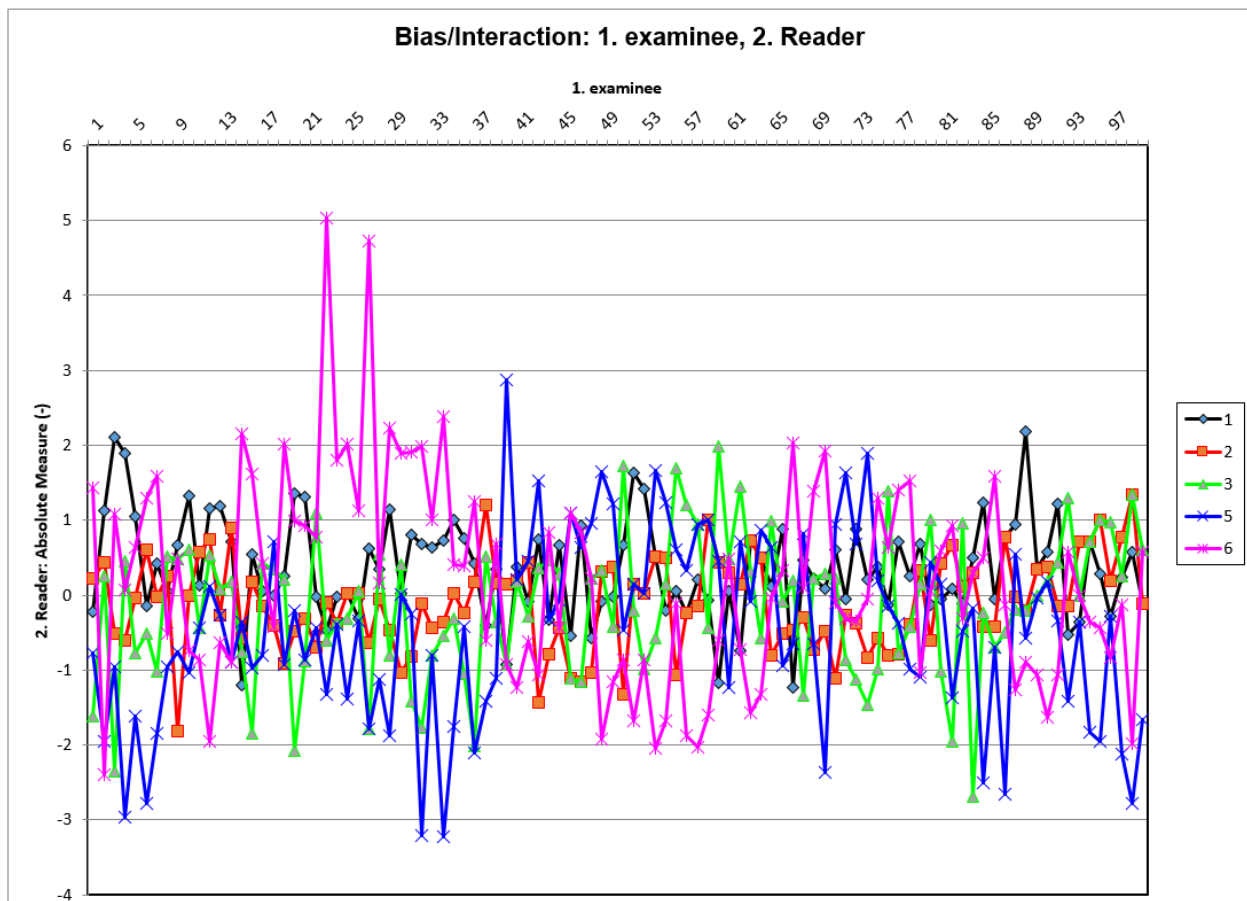
En el caso opuesto, al final de la Tabla 5.5, observamos que el evaluador 6 fue demasiado indulgente con el examinado 12. Le otorgó un total de 23 puntos cuando el puntaje esperado era de 11.93. En la Gráfica 5.4 se aprecia que el evaluador 6 es demasiado laxo.

Gráfica 5.4: Sesgos e interacciones entre evaluadores y el examinado 12



En la Gráfica 5.5 podemos observar las evaluaciones de los 100 examinados y los 5 evaluadores. Esta gráfica también permite evidenciar quiénes son los evaluadores con mayores sesgos en sus evaluaciones. Como se aprecia claramente, los evaluadores 5 y 6 son quienes presentan las evaluaciones más severas e indulgentes de manera asistemática en comparación con los demás evaluadores. Esto se puede apreciar en los picos de color rosa (evaluador 6) y azul (evaluador 5), respectivamente. Esta gráfica es de gran ayuda para identificar textos que están resultando problemáticos para un evaluador, los cuales podrían ser utilizados para dar retroalimentación individual.

Gráfica 5.5: Sesgos e interacciones entre evaluadores y examinados



5.4.3. LA PERCEPCIÓN DE LOS EVALUADORES SOBRE LA RÚBRICA

Una vez que los evaluadores concluyeron la evaluación de los textos, se les pidió que contestaran un cuestionario en línea sobre su experiencia con el uso de la rúbrica y las sesiones de capacitación²⁰. Las primeras cuatro preguntas abordaban aspectos académicos de los evaluadores, mientras que las siguientes se encontraban relacionadas con la rúbrica y el proceso de corrección de los textos.

Los evaluadores eran todos docentes de español como lengua extranjera y con formación académica diversa: doctorado en lingüística, maestría en lingüística, licenciatura en letras hispánicas, pedagogía, filología inglesa y letras, respectivamente. Debido a que la responsable del examen (evaluadora 1) y el investigador (evaluador 2) participaron en el proceso de rediseño de la rúbrica, solamente los evaluadores 3, 4, 5 y 6 contestaron la sección de la encuesta que se relacionaba con la rúbrica y las sesiones de retroalimentación. De estos cuatro evaluadores, solamente la evaluadora 4 expresó tener experiencia en la evaluación de pruebas escritas mediante rúbricas analíticas en el Centro Nacional de Evaluación (CENEVAL). Sin embargo, su experiencia era con exámenes de certificación de natiohablantes del español y con un fin de evaluación distinto.

Con respecto a las preguntas subsecuentes, éstas se describirán de forma individual.

P.5 ¿Cuál es su opinión general sobre la rúbrica?

Los cuatro evaluadores dieron comentarios positivos, expresando que ésta era clara, fácil de entender y apropiada para los fines de la evaluación. Sin embargo, los evaluadores 3, 5 y 6 resaltaron que hacían falta ejemplos de los distintos niveles de desempeño incluidos en las categorías. A esto, la evaluadora 6 expresó que: “No hay ejemplos específicos, por lo que el evaluador requiere de cierta práctica para saber con claridad qué se está evaluando”. La evaluadora 4 también agregó que en vez de contar con 5 páginas de la rúbrica, esta información podría condensarse de forma más visual en tan solo dos páginas.

P.6 ¿Considera usted que hay alguna categoría que debería agregarse o suprimirse?

Los evaluadores 3 y 5 mencionaron que las categorías eran adecuadas y que la rúbrica era precisa. La evaluadora 6 señaló, nuevamente, que hacían falta ejemplos en cada uno de los

²⁰ Los resultados completos del cuestionario se encuentran en el Anexo 2.9.

niveles de desempeño de las categorías. La evaluadora 4 mencionó que en el desarrollo de ideas, sería mejor incluir conceptos tradicionales de evaluación como “insuficiente, suficiente o incipiente”.

P.7 *¿Considera usted que el número de bandas (0 al 3) es suficiente para evaluar cada criterio?*

Los evaluadores 3, 5 y 6 concordaron que sí eran suficientes. Los cuatro evaluadores mencionaron, sin embargo, que en cada categoría debería haber distintas bandas de desempeño, dependiendo de lo que se estaba midiendo.

P.8 *¿Hubo alguna sección de la rúbrica que le pareciera poco clara para evaluar los textos?*

Los evaluadores 5 y 6 respondieron que todo estaba claro. La evaluadora 3 encontró dificultades en identificar a qué se refería el criterio de “léxico especializado”. La evaluadora 4 mencionó que la cohesión y la coherencia se encontraban íntimamente relacionadas, por lo cual la distinción entre ambas no era muy clara en la rúbrica.

P.9 y P.10 *¿Hubo alguna categoría o nivel que le fue particularmente difícil emplear? En caso de haber encontrado dificultades en el uso de alguna categoría o nivel de la rúbrica, ¿qué estrategia o criterio empleó para evaluar dicha categoría o nivel?*

El evaluador 5 dijo que no. La evaluadora 3 respondió con respecto al número de palabras solicitado, puso más énfasis en la información contenida en el escrito que en el número de palabras empleado: “Creo que siempre antepuse el hecho de que se cumpliera el objetivo comunicativo”. La evaluadora 4 mencionó que la coherencia y la cohesión, para subsanar dificultades, empleó criterios personales como: “suficiente e insuficiente”. La evaluadora 6 expresó que, en general, era difícil relacionar los criterios de cada una de las categorías con los textos. La evaluadora 6 mencionó que, en general, en todas las categorías, pues la longitud del texto generaba representaciones mentales distintas de emparejar con cada una de las categorías de la rúbrica.

P.11 *¿Según su criterio, jerarquice las categorías de la rúbrica de mayor a menor importancia?*

Evaluador 3: Adecuación, desarrollo de ideas, cohesión y coherencia, gramática, léxico y ortografía.

Evaluador 4: Gramática, ortografía, léxico, desarrollo de ideas, adecuación, cohesión y coherencia.

Evaluador 5: Logro de tarea, cohesión, coherencia, desarrollo de ideas, gramática, léxico y ortografía.

Evaluador 6: Gramática, ortografía, cohesión, coherencia, léxico, desarrollo de ideas, logro de la tarea.

P.12 *Según su opinión, ¿considera que las 4 temáticas eran equiparables en dificultad?*

La evaluadora 3 dijo que sí, pero que le parecía que el tema de la salud era más inspirador para desarrollar ideas. La evaluadora 4 afirmó que no, razón por la cual se requería de diferentes bandas en las rúbricas. El evaluador 5 reportó que las gráficas eran más difíciles que los ensayos, y que algunos estudiantes se confundieron en su interpretación. La evaluadora 6 dijo que sí.

P.13 *¿Encontró alguna diferencia sustancial al evaluar textos escritos a mano y en computadora?*

Los evaluadores 3, 5 y 6 comentaron que eran más fáciles de evaluar en la computadora porque eran más legibles. La evaluadora 4 mencionó que no había diferencia.

P.14 *Expresar su opinión de acuerdo con la retroalimentación que recibió en el uso de la rúbrica.*

La evaluadora 3 mencionó que fue importante para aclarar dudas y unificar criterios. La evaluadora 4 dijo que no fue clara. El evaluador 5 respondió que fue precisa y le ayudó a utilizar la rúbrica de la mejor manera posible. La evaluadora 6 comentó que tenía una representación distinta de la evaluación, pero que después de la retroalimentación fue más fácil comprender los criterios requeridos para evaluar los textos: “Fue un enfoque diferente y, aprendí algo nuevo, pero, sin duda, es necesaria mayor exposición a este tipo de criterios”.

P.15 *¿Considera usted que hizo falta algo en su proceso de evaluación de los textos?*

Los evaluadores 3 y 5 dijeron que no. La evaluadora 4 afirmó que era necesario señalar que el propósito era evaluar el desempeño de no nativohablantes del español. La evaluadora 6 escribió que hacía falta más capacitación en el uso de la rúbrica.

5.4.4. EL ALCANCE DE LA RÚBRICA DE LA PE PARA EVALUAR HABILIDADES DE ESCRITURA ACADÉMICA

Una vez que se afinó la primera versión de la rúbrica con base en el pilotaje, se llevó a cabo el emparejamiento de la segunda versión con las habilidades de escritura necesarias para que los estudiantes universitarios sean académicamente competentes. Del análisis que se llevó a cabo con cinco especialistas en diseño y evaluación mediante rúbricas para evaluar pruebas

de respuesta construida, se tomó como criterio de selección aquellas habilidades de escritura, evaluadas por la rúbrica, que fueron identificadas por la mayoría de los especialistas; es decir, por lo menos 3 de los 5 especialistas (ver Anexo 2.10). El Cuadro 5.17 muestra el resultado del juicio de expertos. En la columna de la izquierda se aprecian las habilidades de escritura académica que sí mide la rúbrica del EXELEAA, y en la columna de la derecha, aquellas que no mide.

Cuadro 5.17: Resultado del juicio de expertos

Habilidades de escritura que sí mide la rúbrica del EXELEAA	Habilidades de escritura que no mide la rúbrica del EXELEAA
<p>23. Organizar las ideas y la información coherentemente.</p> <p>33. Evitar errores ortográficos.</p> <p>24. Escribir de forma precisa y concisa, evitando frases vagas o vacías.</p> <p>3. Abstractar o resumir información esencial.</p> <p>16. Presentar los datos y la información de manera clara y lógica.</p> <p>6. Analizar y sintetizar información de múltiples fuentes (incluye comparación y contraste).</p> <p>5. Explicar un hecho o suceso utilizando evidencia.</p> <p>32. Utilizar correctamente la gramática del español escrito con el fin de evitar que el lector se distraiga o que se cometan errores que alteren el significado.</p> <p>8. Escribir persuasivamente mediante la construcción de un argumento bien razonado para apoyar o refutar una postura.</p> <p>30. Usar vocabulario técnico y específico al contenido del tema con precisión y de forma apropiada para un propósito y audiencia en particular.</p> <p>10. Examinar el razonamiento en un argumento dado y discutir las fortalezas y debilidades de su lógica.</p> <p>9. Explorar las relaciones entre ideas complejas y posiblemente contradictorias.</p> <p>31. Elegir palabras efectivamente.</p> <p>15. Escribir adecuadamente para un público bien informado y reflexivo.</p>	<p>20. Dar crédito adecuado a las fuentes.</p> <p>29. Revisar y editar el texto para mejorar su calidad, coherencia y precisión.</p> <p>19. Integrar el material citado y referenciado debidamente dentro del texto.</p> <p>26. Escribir con fluidez, evitando que el lenguaje sea complicado y rebuscado.</p> <p>12. Interpretar los datos dentro de un marco determinado.</p> <p>2. Explicar cómo se lleva a cabo un procedimiento.</p> <p>11. Identificar los problemas en una línea de acción propuesta o en la interpretación de hechos y proponer soluciones o alternativas de interpretación.</p> <p>1. Describir observaciones.</p> <p>13. Clasificar la información de acuerdo con categorías o jerarquías.</p> <p>7. Predecir consecuencias o resultados mediante el análisis de información, patrones o procesos.</p> <p>27. Variar la estructura de las oraciones para comunicar ideas de manera efectiva.</p> <p>28. Expresar ideas en forma original o novedosa para mantener el interés del lector.</p> <p>17. Usar la analogía, metáfora o comparación para definir o explicar conceptos técnicos o abstractos para una audiencia general.</p>

<p>21. Desarrollar un debate bien enfocado y debidamente sustentado mediante ejemplos relevantes.</p> <p>25. Escribir con claridad mediante transiciones suaves de un pensamiento a otro.</p> <p>4. Expresar opiniones personales.</p> <p>22. Aclarar las relaciones entre ideas principales y secundarias</p> <p>18. Utilizar formatos eficientes y claros para organizar la información y guiar al lector.</p> <p>14. Utilizar las convenciones de un género en particular.</p>	
---	--

5.5. DISCUSIONES, LIMITACIONES Y FUTURAS INVESTIGACIONES

Las discusiones se presentan con base en las dos inferencias establecidas en este segundo estudio: evaluación y generalización.

5.5.1. INFERENCIA DE EVALUACIÓN

Esta inferencia se respaldó mediante la evidencia recolectada para tres aseveraciones relacionadas con: el estándar de calidad de la rúbrica mediante MFRM, la percepción de los evaluadores sobre la rúbrica y el alcance de la escala de evaluación para medir las habilidades de escritura académica requeridas a nivel posgrado.

5.5.1.1. LA CALIDAD DE LA RÚBRICA SEGÚN LOS PARÁMETROS DE MFRM

Como se pudo apreciar en los mapas de variabilidad y en las tablas arrojadas por FACETS, la segunda versión de la rúbrica no solamente demostró una mayor calidad en comparación con la primera, sino que sus parámetros de calidad resultaron adecuados para la evaluación de las tareas de la PE del EXELEAA. La única categoría que se encontró ligeramente fuera de rango fue la 7 (ortografía). De hecho, la medición de severidad de esta categoría fue de -1.23 lógitos, seguida de la adecuación con 0.00 lógitos. Esto significa que la categoría fue empleada por los evaluadores con suma indulgencia. Lo anterior se podría explicar debido a que el sistema de escritura en español es muy similar a su fonética. A diferencia de otras lenguas que también emplean el alfabeto grecolatino (p.ej., alemán, inglés, francés e italiano), en español, casi cada sonido (excepto por la *h* y algunos casos de consonantes dobles) tiene una representación gráfica. Las consonantes dobles solo aparecen por escrito cuando existe un sonido distinto (*rr*,

ll, cc). Los errores más comunes se presentan con homófonos (presencia y ausencia de la *h*, *v* y *b*, y *c*, *s* y *z*) y acentos ortográficos. Sin embargo, como se pudo observar en los resultados estadísticos, es claro que incluso los estudiantes no hispanohablantes de nivel bajo obtuvieron buenos resultados en la ortografía de la PE del EXELEAA.

Adicional a la indulgencia con que se evaluó la ortografía, ésta fue la única categoría con un ligero desajuste de *Outfit* (1.24) y la única que desplegó un *PtMea* ligeramente menor al predicho por el modelo (0.59). En suma, se puede apreciar que esta categoría aún requiere refinamiento y no ayudó en gran medida a discriminar entre los estudiantes menos proficientes de aquellos más aptos. Una posible solución sería combinar esta categoría con la gramática en un apartado llamado: aspectos mecánicos de escritura. Otra posible solución sería fusionarla con el léxico en una categoría denominada: léxico y ortografía. No obstante, el resto de las categorías presentaron índices de calidad óptimos para la evaluación de los textos.

5.5.1.2. LA PERCEPCIÓN DE LOS EVALUADORES SOBRE LA RÚBRICA

Recordemos que la evaluadora 4 fue eliminada del estudio dado que presentaba índices de desajuste importantes; por lo cual se describirá primero su perfil y posteriormente se mencionarán aspectos generales que destacaron en el cuestionario.

En general, los comentarios de la evaluadora sobre la rúbrica y la capacitación fueron negativos. Por una parte, reportó problemas en el uso de la rúbrica, tanto de formato como de contenido. La evaluadora comentó que los descriptores eran muy extensos, que la nomenclatura de cada uno de los niveles no era muy clara, que el número de bandas podría aumentar en ciertos criterios y que los términos de cohesión y coherencia no eran muy claros. Adicionalmente, comentó que uno de los criterios más importantes debería ser la gramática y ortografía, lo cual la caracteriza con un perfil de enfoque en la microestructura y enfoque en habilidades de orden inferior de escritura. También agregó que las 4 temáticas no eran equiparables en dificultad. Curiosamente, aseveró que el modo de escritura no implicaba ninguna diferencia al momento de evaluar. Por otra parte, en cuanto a la retroalimentación en el uso de la rúbrica, también comentó que no fue muy clara y que sería necesario enfatizar que se trataba de textos en lengua extranjera.

En el caso de los otros tres evaluadores novatos (3, 5 y 6), su opinión general con respecto a la rúbrica fue que era clara, fácil de comprender, pero que era necesario contar con más

ejemplos de cada uno de los niveles de desempeño de las distintas categorías. En relación con el número de bandas, parece que un punto de vista común entre los evaluadores era que cada categoría debería tener distintas bandas. Esto representa un problema al momento de llevar a cabo análisis estadísticos (p.ej., mediante MFRM) puesto que para que las categorías sean comparables entre sí, deben contar con el mismo número de bandas. Cuando las categorías no son claras o difíciles de implementar en un texto en particular, los evaluadores demostraron valerse de estrategias de comparación –por ejemplo, con respecto a otro texto– o de interpretación personal para poder asignarle un puntaje al examinado.

En relación con la jerarquización de la rúbrica, los evaluadores 3 y 5 coincidieron en los aspectos superiores de escritura (logro de la tarea, cohesión, coherencia y desarrollo de ideas), seguidos de los aspectos a nivel microestructura (gramática, léxico y ortografía), mientras que los evaluadores 4 y 6 dieron mayor importancia a aspectos de orden inferior de escritura. Curiosamente, los dos primeros evaluadores reportaron índices de calidad mejores que los dos últimos. Quizás exista una relación entre el perfil cognitivo del evaluador y su habilidad para evaluar textos. Desde luego que esto tendría que confirmarse en estudios con muestras más amplias de evaluadores.

Con respecto a la dificultad de las temáticas propuestas, las opiniones fueron diversas. Parece ser que hay temas que los evaluadores perciben más complejos que otros y también temáticas que favorecen más el desarrollo de las ideas. En relación con el modo de administración, los tres coincidieron en que los escritos en computadora eran más legibles y fáciles de evaluar.

En relación con la retroalimentación, los evaluadores 3, 5 y 6 reconocieron la importancia de estas sesiones para poder clarificar dudas y homogeneizar los criterios de la evaluación de los textos. No obstante, comentaron que hacía falta más entrenamiento, ejemplos y sesiones de retroalimentación. Esto da cuenta de la importancia de contar con un manual de capacitación para los evaluadores, sin importar si se trata de evaluadores novatos o experimentados, que contenga ejemplos detallados en cada una de las categorías. También parece ser que es indispensable un acompañamiento en su proceso de evaluación de textos para unificar los criterios de la evaluación.

5.5.1.3. EL ALCANCE DE LA PE PARA EVALUAR HABILIDADES DE ESCRITURA ACADÉMICA

Como se pudo apreciar en el Cuadro 5.17, resulta notorio que la PE del EXELEAA mide algunas habilidades de escritura académica y otras no. En primer lugar, el examen evalúa la capacidad de escribir textos argumentativos que contienen: una tesis, argumentos y contraargumentos fundamentados mediante alguna operación retórica (secuencia, ejemplificación, definición, comparación o contraste), y una conclusión. En segundo lugar, el examen también mide la descripción y el contraste de la información relevante de gráficas, así como la síntesis de información secundaria y una conclusión que resume y establece la relación entre los datos. Dentro de estos dos grandes grupos, la PE, de forma interna mide: la adecuación del registro en cuanto a las convenciones formales académicas, la cohesión: la relación entre oraciones (recursos de referencia interna) y entre párrafos (nexos sintácticos y signos de puntuación) y la coherencia: la relación semántica entre las ideas y los párrafos (uso de marcadores discursivos), así como el uso de vocabulario académico y de términos especializados, la precisión gramatical y la ortografía.

No obstante, se puede ver claramente que el examen no mide la habilidad del examinado de construir conocimiento a través de múltiples fuentes en las cuales es importante dar crédito a éstas, así como tampoco mide la integración adecuada del material citado y referenciado dentro del texto. De igual manera, el examen tampoco mide el conocimiento previo del tema que permita: interpretar datos o información dentro de un marco determinado, identificar problemas en una línea de acción propuesta, clasificar la información de acuerdo a categorías o jerarquías, predecir consecuencias o resultados mediante el análisis de información. En cualquier caso, la falla no es de la rúbrica per se, sino más bien del constructo subyacente al propósito de la evaluación. En otras palabras, la rúbrica simplemente es resultado de las tareas que se eligen para estimular la escritura de los examinados.

Como se pudo ver en el primer estudio de esta investigación, las habilidades de escritura académica más importantes a nivel posgrado son aquellas relacionadas con la construcción del conocimiento. Por esta razón, exámenes como el TOEFL iBT y Pearson *Academic* han incorporado tareas integradas en sus exámenes. Desafortunadamente, la inclusión de estas tareas también ha traído consigo otras problemáticas. En primer lugar, Cumming (2014) cuestiona si estas tareas en realidad son representativas del constructo de escritura en el “mundo real”. En efecto, los exámenes de lengua con fines académicos deberían incorporar tareas

integradas de lectura para la escritura en donde se puedan evidenciar habilidades de resumen y de parafraseo y síntesis. Sin embargo, la inclusión de estas tareas, como se expuso en el estudio anterior, conlleva dificultades para los evaluadores humanos (i.e., determinar lo que se ha copiado literalmente del texto, distinguir entre lo que puede producir y comprender el examinado). No obstante, pese a las dificultades que entraña la inclusión de tareas integradas, se debería explorar su incorporación en los ECLAA.

5.5.2. INFERENCIA DE GENERALIZACIÓN

Para esta inferencia solamente se enunció una aseveración: los puntajes asignados por los evaluadores alcanzan los estándares de calidad determinados para MFRM. Como se pudo ver en los resultados, la evaluadora 4 tuvo que ser descartada del estudio debido a que sus parámetros de calidad eran asistemáticos y perjudiciales para la evaluación (*Outfit* de 1.64). Una vez que ésta fue removida, se corrió nuevamente el análisis estadístico. En esta ocasión, la evaluadora 6 se encontró ligeramente fuera del rango óptimo de evaluación. No obstante, los resultados de los evaluadores, en general, fueron satisfactorios.

Cabe mencionar que había cuatro evaluadores novatos, por lo que se esperaba que hubiera ciertas discrepancias en las evaluaciones. Como se pudo observar en los ejemplos mostrados, el análisis de MFRM es de gran utilidad para determinar qué evaluador está siendo demasiado severo o indulgente con un examinado en particular. Esto permite ofrecer retroalimentación a los evaluadores con respecto a su desempeño, pero también sirve para identificar textos que resultan problemáticos para evaluar.

El problema de los análisis estadísticos es que no nos permiten comprender por qué razón un evaluador presenta problemas de interpretación de la rúbrica o de los textos. Más aún, tampoco podemos saber qué hacen estos evaluadores para subsanar deficiencias al momento de asignar un puntaje. Por este motivo, el análisis cualitativo a través del cuestionario fue de gran ayuda para comprender qué sucede con los evaluadores cuando corrigen los textos mediante rúbricas analíticas. Por ejemplo, de los resultados estadísticos y cualitativos, se pudo observar que la evaluadora 4 presentó severas dificultades para asimilar cuál era el propósito y criterio para la evaluación. De hecho, esta evaluadora fue convocada por su amplio perfil docente y académico, aunado a su experiencia en evaluaciones de PE para exámenes administrados a gran escala. No obstante, el fin de la evaluación a la que ella estaba acostumbrada era distinto y también en lengua materna. Probablemente esta fue la razón principal por la cual tuvo tantas

dificultades en comprender la rúbrica y ajustarse a la evaluación de textos producidos por no hispanohablantes y con fines académicos. Dentro de sus comentarios vertidos en el cuestionario también se pudo vislumbrar una forma “tradicional” de evaluación enfocada más hacia aspectos mecánicos de la escritura.

5.5.3. LIMITACIONES Y FUTURAS INVESTIGACIONES

Este segundo estudio presenta algunas limitaciones. En primer lugar, lo ideal habría sido que tanto en el pilotaje con la primera versión de la rúbrica, como en la sesión de corrección de los 100 textos con la segunda versión de la escala, se emplearan las mismas temáticas de las tareas, el mismo modo de administración (i.e., en papel o en computadora) y la participación de los mismos evaluadores. Sin embargo, debido a que el estudio se condujo con un examen que estaba emergiendo como alternativa al examen existente, se tuvieron que hacer adecuaciones a las condiciones existentes en el momento. Por ejemplo, en el caso de los evaluadores, habría sido muy enriquecedor conocer también su opinión respecto de las dos versiones de la rúbrica. Lamentablemente, dado que quienes colaboraron en el pilotaje lo hicieron de forma no remunerada, su carga de trabajo no les permitió participar en la segunda sesión de evaluación.

Adicionalmente, habría sido más fructífero contar con un panel de jueces más numeroso. Desafortunadamente, la carencia de docentes universitarios experimentados en el diseño y uso de rúbricas para evaluar escritos en ECLAA imposibilitó que se reclutara un número mayor de éstos. Esta sesión se llevó a cabo de forma individual, pero también habría sido valioso contar con una sesión conjunta en la cual los jueces externaran las razones por las cuales consideraban que las habilidades de escritura académica importantes para el posgrado eran evaluadas o no por la rúbrica. Si bien estas sesiones habrían permitido llegar a un consenso común, también es cierto que se corre el riesgo de que los jueces se vean sesgados por la visión de los demás, con lo cual podrían cambiar su veredicto no por estar realmente convencidos, sino por la influencia que en ocasiones puede ejercer la visión de uno de ellos.

En cuanto a la inferencia de evaluación, en futuras investigaciones se podría hacer una comparación entre la calidad en la evaluación mediante una rúbrica analítica y una holística. Desafortunadamente, los estudios que se han realizado en este sentido resultan aún contradictorios. Por una parte, hay quienes han encontrado que la evaluación es más confiable y de mayor calidad cuando se usan escalas holísticas, pero también hay quienes arguyen lo opuesto. En este mismo sentido, los cuestionarios o entrevistas que se administran a los

evaluadores también podrían dar cuenta de qué escala les parece más fácil de emplear para evaluar los textos.

En este estudio no se incorporó ninguna entrevista, pero sería de gran ayuda, posterior a la administración del cuestionario, entrevistar a los evaluadores para ahondar en aspectos relevantes que hayan surgido al momento de responderlo. Adicionalmente, los cuestionarios y entrevistas también serían de gran ayuda para evidenciar si los evaluadores prefieren el uso de rúbricas analíticas u holísticas, y cómo estos resultados cualitativos se relacionan con su calidad evaluando con una u otra escala.

Finalmente, valdría la pena llevar a cabo estudios de entrevista rememorada justo cuando los evaluadores han corregido un texto. Este tipo de estudios ayudaría significativamente a comprender qué aspectos están tomando en cuenta los evaluadores al momento de evaluar los textos. De igual forma, esto ayudaría a comprender si el proceso cognitivo que siguen cuando emplean una rúbrica analítica u holística difiere.

5.6. CONCLUSIONES

La validez interna de la PE de un ECLAA generalmente se evidencia mediante la calidad de la rúbrica y de los puntajes que otorgan los evaluadores. Sin embargo, un aspecto que generalmente se deja del lado es la extensión de la rúbrica al contexto para el cual deseamos evaluar. En este sentido, en este estudio se abordó tanto la calidad del proceso de evaluación (i.e., las tareas, los evaluadores y la rúbrica) como la factibilidad de la rúbrica de medir las habilidades de escritura académica importantes a nivel posgrado. En cuanto a la calidad del proceso de evaluación, los análisis estadísticos mediante MFRM demostraron que la diferencia de la dificultad de las tareas era mínima, la segunda versión de la rúbrica demostró ser de mejor calidad que la primera –aunque todavía la categoría de la ortografía requería refinamiento– y los evaluadores también desplegaron un alto grado de consistencia y confiabilidad en sus resultados (excepto por uno de ellos que tuvo que ser descartado del análisis). En relación con el cuestionario, se pudo percibir que la rúbrica, en general, fue clara y de utilidad para el proceso de corrección. El proceso de capacitación y de retroalimentación fue también de suma importancia para establecer criterios comunes del proceso de evaluación y para clarificar dudas, aunque se destacó que hacían falta más ejemplos en cada categoría y nivel de desempeño, así como también más capacitación. Finalmente, con respecto al alcance de la rúbrica para evaluar

las habilidades de escritura académicas, se evidenció que ésta evalúa las habilidades relacionadas con aspectos pragmáticos, sociolingüísticos y lingüísticos, pero no evalúa habilidades relacionadas con la construcción del conocimiento ni la incorporación de conocimiento previo. Por esta razón, valdría la pena considerar la inclusión de tareas de respuesta construida en un ECLAA.

CAPÍTULO 6

ESTUDIO 3: LA EXTRAPOLACIÓN DE LOS RESULTADOS DE LA PE DEL EXELEAA AL “MUNDO REAL”

6.1. INTRODUCCIÓN

El fin de la evaluación de alto impacto no culmina en el resultado que se le otorga al examinado. Es cierto que el examen y el proceso de corrección deben ser válidos, confiables, justos y equitativos; no obstante, este tipo de evaluación conlleva importantes implicaciones en la vida del examinado. Por este motivo, los exámenes de alto impacto deben ser evaluados tanto en su validez interna como externa. Es decir, la validez de un examen no se circunscribe exclusivamente a qué tan bien funciona el instrumento de evaluación, sino también a cómo se puede extrapolar éste al “mundo real”.

En este sentido, la cuarta y última inferencia de la presente investigación se relaciona con la validez externa del examen; es decir, la extrapolación. En los exámenes de alto impacto, uno de los escollos radica en lo difícil que es extrapolar los resultados del examen al mundo real. Esto debido a lo compleja e intrincada que resulta la tarea de relacionar lo que un sustentante demuestra, lingüísticamente hablando, en un examen (p.ej., en la PE del EXELEAA) y las prácticas de escritura en su quehacer cotidiano para demostrar la adquisición de conocimiento.

Esta inferencia presupone dos cosas: por una parte, que los resultados de la PE correlacionan de forma positiva con otros instrumentos propios del ámbito para el cual se desea evaluar –en este caso, del contexto académico–. Por otra parte, se infiere que un examinado que demuestra buen dominio de las habilidades de escritura evaluadas en la PE del EXELEAA, por ejemplo, será capaz de escribir eficientemente dentro de una amplia gama de tareas y trabajos académicos que le serán requeridos durante sus estudios universitarios. Contrariamente, también debería confirmarse lo opuesto; es decir, que quienes obtienen un puntaje bajo en la PE, también presentan deficiencias y tropiezos para escribir apropiadamente en el contexto universitario. De ahí que, resulte crucial respaldar mediante evidencia la inferencia de extrapolación dentro del proceso de validación de un examen de alto impacto.

6.2. MARCO TEÓRICO

La extrapolación representa un tipo de validez externa, pues se toman decisiones con base en lo que se cree que el examinado será capaz de desempeñar en el “mundo real”. En el caso de los exámenes de certificación, es importante conducir estudios en este sector para poder dar cuenta de que un sustentante que aprueba un examen (en este caso de lengua con fines académicos) será capaz de hacer uso de la lengua en el contexto para el cual se lo evalúa. Los estudios de extrapolación más socorridos han tratado de correlacionar el puntaje que se obtiene en un ECLAA con las notas académicas que obtienen los estudiantes universitarios. Recientemente, los estudios de extrapolación se han realizado mediante comparaciones textuales entre el escrito producido en un ECLAA y los trabajos académicos de los alumnos. Finalmente, debido a la escasa correlación que estos dos tipos de estudios han presentado, actualmente se están buscando formas alternas que permitan dar cuenta de cómo el resultado de un examen se puede extrapolar al “mundo real” (p.ej. a través de entrevistas o cuestionarios). En este apartado se detallarán cada una de estas aproximaciones a la inferencia de extrapolación.

6.2.1. LA CORRELACIÓN DE UN EXAMEN DE LENGUA CON LAS NOTAS ACADÉMICAS

Los estudios de extrapolación de los resultados de un ECLAA al “mundo real” que mayor auge han tenido son aquellos que pretenden correlacionar el puntaje que obtienen los examinados en un ECLAA con las calificaciones que obtienen en sus estudios universitarios. A este tipo de estudios se lo conoce también como validez predictiva, pues uno de los fines de los exámenes de certificación consiste en predecir, con base en el resultado del candidato, el futuro desempeño en situaciones comunicativas similares a las que se evaluaron en el examen. Sin embargo, el problema que se ha suscitado con este tipo de estudios es múltiple. Primeramente, una de las mayores deficiencias de estos estudios reside en la falta de poder evidenciar si los estudiantes que no poseen el puntaje mínimo considerado para admisión podrían haber sido académicamente competentes o no (Cho y Bridgeman, 2012; Fox, 2004). Por ejemplo, Cotton y Conrow (1998) encontraron que un puntaje superior a 7.0 en el IELTS no necesariamente era garantía de éxito académico, ni que tampoco uno menor a 5.5 incuestionablemente representaría un predictor de fracaso académico. La otra problemática que se resalta en este tipo de estudios radica en lo contradictorios que han resultado. Por ejemplo, estudios como los de Feast (2002), Ushioda y Harsch (2011) y Woodrow (2006)

evidenciaron una correlación¹ positiva, entre fuerte y moderada, entre el nivel de dominio de inglés y el desempeño académico de los estudiantes. Otros, sin embargo, han encontrado que estas correlaciones son muy pequeñas o inexistentes (Avdi, 2011; Cotton y Conrow, 1998; Craven, 2012). Finalmente, cabe mencionar que ninguno de los estudios antes mencionados toma en cuenta el fenómeno de atrición² (Kokhan, 2012, 2013); es decir, el nivel de lengua de un sustentante cuando presenta un ECLAA (p.ej., TOEFL o IELTS) es muy distinto al nivel de lengua que posee cuando escribe sus primeros trabajos académicos. De hecho, desde el momento en que un sustentante presenta un ECLAA y el momento en el que recibe las calificaciones de su primer semestre académico, puede transcurrir un lapso de casi tres años. Consecuentemente, ¿qué tan fiables son los estudios de validez predictiva que comparan el nivel de lengua de un sustentante después de más de dos años? En todo caso, estos estudios deberían llevarse a cabo de forma concurrente.

Por ejemplo, en el estudio de Avdi (2011), la autora dividió a un grupo de estudiantes de maestría según su puntaje obtenido en el IELTS. El puntaje del examen oscila entre 0 a 9 puntos, y normalmente el mínimo que se considera como aceptable para admisión es 6.5; sin embargo, dependiendo de la universidad receptora, es posible que acepten a candidatos con un puntaje de 5 siempre y cuando se comprometan a tomar cursos de lengua. La clasificación se hizo en tres bloques: estudiantes con nivel bajo (entre 5.0 y 6.0), estudiantes intermedios (6.5) y avanzados (7.0 y 8.0). Posteriormente, la autora correlacionó los puntajes con los resultados académicos obtenidos por los estudiantes. La correlación de estos puntajes fue baja (correlaciones de 0.359). De manera más puntual, se evidenció que los estudiantes del bloque intermedio (resultados de 6.5 en el IELTS) obtuvieron notas más bajas que los del bloque bajo. Consecuentemente, Avdi (2011), Craven (2012), Fox (2004) y Humphreys, Haugh, Frenton-Smith, Lobo, Michael y Walkinshaw (2012) concluyen que la edad, el tiempo de vivir en el país

¹ Se considera una correlación positiva con valores entre 0 y 1. Si los valores van de 0 a -1, entonces la correlación es negativa. La correlación de 0 es nula, en rangos de 0.01 a ± 0.19 , se considera que la relación es muy baja o despreciable. Entre ± 0.20 y ± 0.39 , la correlación es baja; de ± 0.4 a ± 0.69 , la relación es moderada; de ± 0.70 a ± 0.89 la correlación es fuerte; de ± 0.90 a ± 0.99 ; y de 1 la correlación es perfecta (Samuel y Okey, 2015, p. 24).

² La atrición hace referencia a la pérdida de las habilidades lingüísticas, en este caso de una segunda lengua, a partir del momento en el cual un candidato sustentó un examen para el ámbito académico, hasta que el estudiante es aceptado por una universidad y comienza a emplear sus habilidades lingüísticas. Según Kokhan (2012), en un año las habilidades lingüísticas de un individuo en este escenario varían considerablemente. En consecuencia, las habilidades demostradas en un examen no se pueden equiparar de manera asincrónica, sino que, para sustentar el argumento de extrapolación, éstas deben ser medidas de forma sincrónica.

con la lengua meta, el ajuste cultural, las habilidades interculturales, el uso del idioma fuera de clase, el nivel de motivación, el compromiso, el antecedente cultural y educativo del estudiante, así como su nivel de dominio lingüístico, contribuyen al desempeño académico de los estudiantes.

Uno de los estudios más recientes es el de Cho y Bridgeman (2012). En este estudio, los autores investigaron la correlación entre diversas pruebas de admisión (p.ej., TOEFL iBT, GRE, GMAT y el *Scholarship Aptitude Test*, SAT) y el desempeño académico de 2594 estudiantes de grado y de posgrado pertenecientes a 10 universidades estadounidenses. Los autores no encontraron una fuerte correlación entre el TOEFL iBT y los resultados académicos de los estudiantes ($r_w=0.16$, el cual solamente explica el 3 por ciento de la varianza de las calificaciones obtenidas en el ciclo universitario); sin embargo, los autores arguyen que los patrones de las gráficas demostraron que, en comparación con los demás exámenes antes mencionados, el TOEFL iBT mostraba mejores tendencias predictivas del desempeño de los estudiantes (p. 421). Por ende, los autores concluyen que aunque la correlación es prácticamente nula, ésta podría ser significativa.

Sin embargo, pareciera que este tipo de estudios busca a toda costa encontrar argumentos que le den sustento a la validez predictiva de los ECLAA. El problema de mezclar numerosas variables dentro de estas correlaciones pone en tela de juicio la pertinencia epistemológica de conducir investigaciones de este tipo. De hecho, los autores en su artículo expresan las desventajas para conducir este tipo de estudios (p.ej., la falta de datos adecuados para poder establecer las correlaciones, el fenómeno de atrición, pero quizás lo más riesgoso es que los mismos autores expresan que no existe una correspondencia directa entre la lengua y el desempeño académico).

... a fundamental issue is that there is no one-to-one correspondence between language and academic performance. Language is a crucial factor in learning, but it is only one of many factors. Evidence and common sense suggest that lack of proficiency may interfere with NNES³ students' academic ability when English is a medium of instruction. However, being linguistically proficient does not guarantee academic success. There are many other factors such as motivation, learning strategies, and quantitative skills that also contribute to one's academic performance. One obvious example that attests to this is that not all native speakers are academically successful. By the same logic, there is no reason to expect that all NNES students with high levels of English proficiency would necessarily have high GPAs. In this regard, showing that language does impact academic performance is quite challenging (Cho y Bridgeman, 2012, p. 424).

³ NNES (Non-native English speakers: hablantes no nativos del inglés).

En realidad, una de las razones por las cuales los estudios de validez predictiva han buscado correlacionar el puntaje de un examen de lengua con las calificaciones obtenidas por un estudiante universitario se debe llanamente a que ambas son fáciles de obtener y correlacionar. La pregunta que subyace es si éste debería ser el criterio para conducir este tipo de estudios. Sin lugar a dudas, la carencia de congruencia es una de las razones por las cuales los resultados que se han obtenido en estos estudios son tan diversos y contradictorios.

Siguiendo la línea de Cho y Bridgeman (2012) y Fox (2004), si el factor de la lengua fuera un determinante para el éxito académico, por añadidura, todos los hablantes nativos del español, por ejemplo, serían académicamente exitosos al cursar estudios universitarios en español. De manera más puntual, Ingram y Bayliss (2007) afirman que no es sorprendente que las correlaciones entre el puntaje de un ECLAA y las notas académicas sean bajas, pues un ECLAA predice el comportamiento lingüístico, no el desempeño académico. Ahora bien, otra de las disyuntivas que ha surgido entre dicha correlación se evidencia en los exámenes de lengua que sustentan los hablantes nativos. Es decir, tampoco los hablantes nativos presentan perfiles uniformes de dominio lingüístico, incluso en lengua materna. Por ejemplo, el estudio conducido por González et al. (2014) demuestra claramente que los estudiantes hispanohablantes universitarios de diversas instituciones de educación superior de la Ciudad de México también presentan perfiles disímiles de habilidades lingüísticas en español como lengua materna. Curiosamente, en el estudio de González et al., aunque la gran mayoría de los estudiantes (63.4%) reportó que la clase de español era una de las materias más fáciles del bachillerato, y que su actitud hacia el aprendizaje de la lengua era favorable (47.7%), sus resultados en el Examen de Habilidades Lingüísticas (EXHALING) arrojaron resultados bajos. De hecho, el 16.2% de un total de 4351 estudiantes evaluados, demostró poseer un nivel deficiente de habilidades lingüísticas. Siguiendo los parámetros de calidad de un ECLAA, estos alumnos no serían aceptados a cursar estudios de grado en ninguna universidad. No obstante, cuando se trata de lengua materna, pareciera que dichas habilidades no son importantes dentro del proceso de selección de los candidatos universitarios de grado o de posgrado (por lo menos en el mundo hispanohablante).

6.2.2. LAS COMPARACIONES TEXTUALES

Existen otros estudios que han buscado comparar los textos que produce un estudiante en un ECLAA con los textos que produce en su quehacer académico con el fin de encontrar similitudes. Primeramente, como en el caso anterior, un problema de estos estudios se debe al

fenómeno de atrición, pues las comparaciones textuales distan mucho de un estudiante a otro, dependiendo del tipo de seguimiento que éstos hayan tenido en su aprendizaje de escritura del idioma meta desde el momento de presentar el examen hasta su ingreso a la universidad. Por ejemplo, Paul (2004) condujo un estudio de análisis discursivo para poder evaluar hasta qué punto el desempeño de los estudiantes en el IELTS podría predecir su uso de la lengua en distintas tareas académicas. En su estudio de caso, seleccionó a cuatro estudiantes que sustentaron el IELTS y les dio seguimiento académico a lo largo de un año. Debido a que la investigadora incluyó la parte oral y escrita, se grabaron y transcribieron las interacciones de los estudiantes a lo largo de su primer semestre académico. Asimismo, se recolectaron los trabajos escritos durante este semestre. De forma simultánea, se llevaron a cabo observaciones y entrevistas. Para el análisis de los textos, siguiendo la perspectiva sistémico-funcional (Halliday, 1994; Martin y Rose, 2003), se contemplaron los siguientes criterios para el análisis de la complejidad lingüística, léxica y gramatical: número de sustantivos (simples y complejos), tipo de proceso (comportamiento, acción, mental, relacional, verbal) y circunstancias (tiempo, lugar, modo, significado, bajo qué condiciones y para qué propósito). También se evaluó la postura (argumentación), la organización, la cohesión y la precisión (errores sintácticos, de vocabulario, de cohesión y casos no inteligibles).

Sin embargo, aun cuando la autora afirma que la producción de los cuatro participantes fue muy similar en los textos producidos en el IELTS y en los textos académicos analizados, no se proporciona ningún tipo de correlación estadística que confirme dicho supuesto. Aunado a lo anterior, Paul (2004) menciona que la producción a un nivel micro no necesariamente es un indicador del nivel macro; en otras palabras, el análisis oracional no da cuenta de que la consecución de la tarea escrita (nivel macro) sea satisfactoria o no. Por ejemplo, la organización retórica dista considerablemente de un género discursivo a otro, además de que la postura y el vocabulario, también juegan un rol indispensable para poder redactar un texto perteneciente a una tipología textual particular. En esta misma línea, Chan, Wu y Weir (2013) condujeron un estudio de análisis comparativo textual entre el *General English Proficiency Test* (GEPT⁴) y los textos académicos que producen los estudiantes en cursos universitarios en el Reino Unido.

⁴ El GEPT avanzado es un examen de inglés con fines académicos que se administra en Taiwán para los estudiantes que desean ingresar al ciclo universitario en ese país o en otros. El estudio de Chan et al. (2013) se realizó con la primera PE del examen. Esta tarea es integrada y se le pide al sustentante que escriba un ensayo de aproximadamente 250 palabras a partir de dos textos que se ofrecen como input.

Para poder hacer una comparación textual, los autores emplearon dos métodos: juicio de expertos y también el análisis textual automatizado. Para los análisis automatizados, se emplearon dos programas: *vocabprofile* y *cohomatrix*. En estos dos programas se compararon tres aspectos: léxico, sintácticos y de cohesión entre los textos del ECLAA y de textos escritos dentro del ámbito académico. De forma más puntual, dentro de los aspectos léxicos, se compararon: palabras de alta frecuencia, palabras académicas, palabras de baja frecuencia y promedio de sílabas por palabra. Para el análisis sintáctico, se compararon: el promedio de palabras por oración, la similitud sintáctica entre oraciones, la media del número de modificadores por frases sustantivas, la media del número de palabras antes del verbo principal y la incidencia de operadores lógicos. Finalmente, para la cohesión, se contrastaron en oraciones adyacentes: el traslape de argumentos, el traslape de palabras con raíces similares, el traslape de palabras de contenido, la proporción de referentes anafóricos y la similitud conceptual (Chan et al., 2013, p. 12). En el caso del proceso cognitivo, los autores incluyeron los siguientes parámetros: 1) representación de la tarea y macroplaneación, 2) revisión de la macroplaneación, 3) conexión y generación, 4) selección de ideas relevantes, 5) minuciosa lectura global, 6) organización de ideas en relación con los textos de entrada, 7) organización de ideas en relación con el propio texto del escritor, 8) nivel de edición bajo durante la escritura, 9) nivel bajo de edición tras la escritura, 10) nivel alto de edición durante la escritura, 11) nivel alto de edición tras la escritura (Chan et al., 2013, p. iii). Tras hacer una minuciosa comparación textual y de los procesos cognitivos que emplean los estudiantes para escribir en el ámbito académico y en la PE del GEPT, los autores aseveran que la PE del ECLAA promueve los mismos procesos cognitivos de escritura que las tareas escritas que se asignan mundo real.

Desafortunadamente, una de las deficiencias al conducir estudios para comparar los textos provenientes de un examen y de una tarea académica se encuentra en la naturaleza diversa de la escritura. Por ejemplo, en la discusión sobre el modo discursivo de las tareas de los ECLAA (ver Capítulo 4, Estudio 1), resulta claro que el modo discursivo en un texto escrito para un ECLAA es intrínsecamente diferente al de un texto escrito dentro del contexto universitario. Dentro del ámbito académico, las tipologías textuales son muy variadas y diversas, además de que la redacción de un texto no solamente depende de las habilidades de un sustentante, sino también del conocimiento previo sobre el tema y de las estrategias de las cuales se vale tanto para construir el conocimiento como para revisar y editar su texto. Adicionalmente, el proceso de planeación, desarrollo y edición de un escrito de 350 palabras para un ECLAA será muy distinto a uno académico en el cual se requiere de mayor tiempo para

su redacción. Sin mencionar, claro está, el apoyo del cual puedan valerse los estudiantes no nativohablantes para revisar y editar sus trabajos académicos: correctores automáticos de los procesadores textuales, diccionarios monolingües, bilingües, tesauros, foros de discusión, así como el apoyo de nativohablantes (p.ej., compañeros de clase, amistades, tutores de tesis o profesionales) a quienes los estudiantes recurren para poder editar sus textos.

Por ejemplo, en un estudio reciente sobre comparaciones textuales, Weigle (2015) cotejó aspectos lingüísticos de los ensayos escritos para un ECLAA, tanto por nativo como por no nativohablantes del inglés, con el corpus obtenido de los trabajos académicos (pertenecientes a numerosas disciplinas académicas) escritos por estudiantes que habían obtenido excelentes notas en sus trabajos. El corpus que se generó a partir de los trabajos escritos de los estudiantes (véase Hardy y Römer, 2013) incluyó todo tipo de textos procedentes de posgrado y del último año de la licenciatura, provenientes de cualquier área del saber: Humanidades y Artes, Ciencias Sociales, Ciencias Biológicas y de la Salud, y Ciencias Físico-Matemáticas. A diferencia de los estudios de Paul (2004) y de Chan et al. (2013), la autora encontró que los ensayos de los ECLAA diferían sustancialmente de los textos escritos en el ámbito académico, particularmente en el área de ciencias naturales y de la salud.

En primer lugar, la investigadora encontró una diferencia sustancial entre los patrones lingüísticos del corpus académico y del corpus generado a partir de los textos escritos para el ECLAA. Los ensayos del examen se caracterizaron por ser más narrativos, de opinión y con un discurso dependiente de la situación. En contraste, en el corpus académico se registraron más aspectos lingüísticos que se relacionan con procedimientos o con escritura informativa, particularmente en las ciencias naturales (Weigle, 2015, p. 34). La investigadora sólo encontró algunas similitudes en aspectos lingüísticos entre los textos del ECLAA y el corpus académico en el área de Humanidades y Artes. Estas semejanzas se encuentran únicamente en la expresión de opiniones y de una escritura evaluativa dependiente de la situación.

Weigle (2015) concluye que los textos provenientes de un ECLAA ofrecen aspectos léxicos y gramaticales disímiles a los que se observan en áreas disciplinarias académicas, especialmente fuera de las humanidades. Esto nuevamente demarca las limitaciones que presentan los ECLAA al tratar de generalizar para un ámbito tan amplio como el académico. La autora sugiere que se incluyan tareas integradas en los ECLAA, pues este tipo de tarea es más representativo de la escritura académica que las discretas. Uno de los grandes beneficios de la

PE de los ECLAA, según Weigle, radica en la utilidad de los resultados para ubicar a los prospectos universitarios –tanto nativo como no nativo hablantes– en diversos niveles de cursos de escritura académica. Esta visión hace eco a la propuesta de Fox (2004) de emplear los resultados de los ECLAA no como métodos de exclusión ni de filtro de acceso a las universidades, sino más bien para identificar tempranamente el tipo de ayuda que requieren los estudiantes que, académicamente hablando son excelentes prospectos para cursar estudios universitarios, pese a su nivel de dominio de la lengua meta.

... A timed impromptu essay is limited in terms of its generalizability to other academic genres. However, ... essay tests may be useful in determining whether students have the basic skills needed to move on to disciplinary writing. Thus, the use of such tests for university admission or placement into writing courses may be more justifiable than their use as end-of-course tests, particularly if courses focus on multi-draft writing and writing from sources (Weigle, 2015, p. 37).

6.2.3. LA CORRELACIÓN DEL EXAMEN CON CRITERIOS ALTERNOS: ENTREVISTAS Y CUESTIONARIOS

La falta de congruencia epistemológica al correlacionar los puntajes de un ECLAA con las calificaciones que obtienen los estudiantes universitarios, o al conducir un análisis textual contrastivo entre un texto redactado para un ECLAA y un texto académico, ha puesto en entredicho los estudios de validez predictiva. Por ende, debido a lo complejo que resulta extrapolar los resultados de los exámenes al mundo real, recientemente se han buscado métodos alternos que permitan evidenciar si la habilidad de escritura de un estudiante correlaciona positivamente con otros instrumentos. Por ejemplo, algunos de estos estudios han incorporado cuestionarios de autoevaluación o de evaluación docente sobre las habilidades de escritura de los estudiantes (Fox, 2004; Powers, 2015; Powers, Kim y Weng, 2010; Sinharay, Powers, Feng, Saldivia, Giunta, Simpson y Weng, 2009; Weigle, 2010;). Cho y Bridgeman (2012) critican este tipo de estudios debido a la subjetividad que subyace en este tipo de enfoques, sobre todo, cuando se trata de las evaluaciones de los docentes. Sin embargo, pese a que existe un alto grado de subjetividad y también disparidad en las evaluaciones que emiten los docentes, este tipo de estudios resulta más acertado que los estudios antes descritos, en los cuales múltiples variables ajenas al factor lingüístico inciden en el éxito académico.

... grade point average (GPA) as a measure... is dependent on a number of other factors that language tests do not measure, such as social networks of support, determination, financial security, time available for study, and time spent studying. Predictive validity studies are particularly weak if they offer evidence only of a correlation between language test scores and GPA... (Fox, 2004, p. 442).

Otro argumento de Cho y Bridgeman (2012), ampliamente discutible, es que la validez normalmente se evidencia en término de correlaciones puramente cuantitativas. No obstante, de acuerdo con Kane (2006, 2013), el proceso de validación de un examen, siguiendo la línea argumentativa de Toulmin (1958), debe echar mano de cualquier tipo de evidencia que permita dar sustento y respaldo a las interpretaciones que se establecen de los exámenes (sobre todo aquellos de alto impacto). Entonces, más que pensar en términos de correlaciones y números, se debería abordar la validez tomando en cuenta cualquier tipo de evidencia que permita justificar las interpretaciones y el uso que se hace de los exámenes.

Un claro ejemplo del esfuerzo por conducir este tipo de estudios es el de Fox (2004). En su estudio, la autora empleó técnicas mixtas para poder evidenciar la utilidad del *Canadian Academic English Language* (CAEL) en la predicción del desempeño académico de los estudiantes universitarios de una universidad canadiense. La autora partió de la premisa que existe, lingüísticamente hablando, un umbral mínimo para que los estudiantes puedan ser académicamente competentes. Sin embargo, el aspecto comunicativo, por sí mismo, no es garantía del éxito académico. A diferencia de otros ECLAA administrados a gran escala, el fin del CAEL no es determinar quién accede a la educación superior, sino más bien identificar quién necesita apoyo de inglés académico en sus estudios y a qué grado. El examen está diseñado en un rango de 0 a 90 puntos. Quienes reciben un puntaje inferior a 30 son considerados con un nivel no apto para el ámbito académico y en riesgo potencial de fracaso académico. Aquellos que se encuentran dentro del rango de 40 a 60 puntos, son aceptados para cursar estudios universitarios, pero en combinación con cursos de inglés con fines académicos. Entre menor sea el puntaje en el CAEL, más serán los cursos de inglés que deban tomar y menos los cursos académicos que pueden registrar. Entre mayor sea el puntaje, la situación anterior será inversa. Finalmente, los que obtienen un puntaje mayor a 70 puntos se encuentran exentos de cursos de inglés y pueden llevar toda la carga académica de un nativohablante. En caso de que un alumno que haya sido asignado a cursos de inglés académico abandone sus cursos de lengua, es dado de baja de sus cursos académicos.

La ventaja del estudio de Fox fue que contempló el desempeño académico de los estudiantes con niveles lingüísticos por debajo de lo que sería considerado admisible. Consecuentemente, la primera interrogante sobre el supuesto fracaso académico de un estudiante que es considerado “no apto en términos lingüísticos” desaparece. La autora contrastó los resultados obtenidos en el CAEL con cuestionarios y entrevistas a profesores,

tanto de cursos académicos, como de aquellos que imparten cursos de inglés académico. Asimismo, la investigadora les dio seguimiento académico a los estudiantes a lo largo de dos años, sobre todo a aquellos en riesgo potencial de fracaso académico debido a su bajo nivel de lengua. Los hallazgos más significativos se resumen a continuación. En primer lugar, 16 profesores de inglés con fines académicos identificaron a un total de 24 estudiantes (de un total de 341) como mal ubicados en sus clases. Es decir, según su criterio, el CAEL los ubicó erróneamente en un curso de lengua más elevado. Estos 24 estudiantes fueron identificados con riesgo potencial de fracaso académico debido a que presentaron un nivel de lengua por debajo del umbral. Sin embargo, la mitad de ellos obtuvo buenas notas en sus estudios universitarios y también en sus cursos de inglés con fines académicos (de hecho, cuatro de ellos obtuvieron excelentes calificaciones). Adicionalmente, la autora destaca que estos estudiantes exitosos demostraron un alto índice de asistencia a clase y de participación en la entrega de trabajos tanto en sus cursos de lengua como académicos. De manera contraria, los doce estudiantes restantes que obtuvieron notas bajas demostraron patrones diversos de asistencia; desde muy baja hasta muy alta. No obstante, los seis que presentaron menor frecuencia de asistencia abandonaron sus cursos durante el primer año académico.

Entonces, Fox (2004) sugiere que la asistencia y la motivación son factores importantes para el éxito académico. De hecho, tras dos años de seguimiento, los cuatro estudiantes con bajo nivel de inglés, pero con alto desempeño académico, fueron descritos como “activos, efectivos o productivos”. En conclusión, la autora destaca que aunque es cierto que los estudiantes con un nivel de lengua por debajo del umbral corren un alto riesgo de abandonar sus estudios, también es cierto que no siempre es así. Asimismo, con base en su estudio, Fox (2004) especula que el proceso de aculturación, la actitud y la motivación, reflejada en la asistencia a clase, la participación y el cumplimiento de tareas, también juegan un papel importante en el desarrollo de habilidades comunicativas y en el éxito académico.

De igual forma, la percepción de los docentes también es un indicador del éxito de los estudiantes tanto en los cursos de inglés con fines académicos como en los cursos curriculares. A este respecto, la autora clasificó los comentarios de los docentes en seis categorías que emergieron de las entrevistas, cuestionarios e intercambio de correos electrónicos; a saber: la participación y preparación, la calidad de los trabajos, la habilidad lingüística del estudiante, la percepción del estudiante sobre la necesidad de recibir apoyo con cursos de inglés con fines académicos, la calidad del conocimiento dentro del área de estudio,

y el entendimiento de los requerimientos o expectativas de los trabajos académicos. Una de las conclusiones más sobresalientes del estudio de Fox (2004) fue que la intervención temprana por parte de los docentes podría prevenir el fracaso de los estudiantes que se encuentran en riesgo por poseer un nivel de lengua inferior al umbral.

Otros estudios recientes sobre validez predictiva (Powers 2015; Powers et al. 2010, Sinharay et al. 2009) corresponden al *Test of English for International Communication* (TOEIC). Si bien este examen se utiliza con fines laborales en el área de negocios, los instrumentos que se emplean resultan extrapolables a los ECLAA. Los investigadores incorporaron cuestionarios de autoevaluación y percepción docente sobre el desempeño de sus estudiantes con el fin de poder contrastar lo que los examinados pueden realizar en situaciones de la vida real.

Los estudios de Powers et al. (2010) y Sinharay et al. (2009) buscaron establecer la correlación de los puntajes obtenidos en el TOEIC y aquellos obtenidos mediante cuestionarios de autoevaluación de los estudiantes. En estos estudios, los investigadores encontraron correlaciones moderadas y altas entre los puntajes obtenidos en el examen y el cuestionario. En el caso del estudio de Sinharay et al. (2009), adicionalmente a los cuestionarios de autoevaluación, los investigadores también diseñaron un cuestionario para que los docentes evaluaran el desempeño de sus estudiantes. También en este contraste de resultados, los investigadores encontraron correlaciones moderadas entre las evaluaciones emitidas por los docentes y los resultados arrojados por el examen.

En un estudio más reciente, Powers (2015) buscó identificar si existía una correlación positiva entre el puntaje obtenido en las cuatro secciones del TOEIC y el puntaje obtenido en un cuestionario de autoevaluación que contenía 24 descriptores. En el estudio participaron estudiantes coreanos y japoneses, quienes mediante un cuestionario en escala tipo Likert del 1 al 5 contestaron si consideraban que podían realizar determinadas tareas comunicativas. El autor encontró correlaciones moderadas entre el examen y el cuestionario de autoevaluación de habilidades comunicativas.

Sin embargo, los resultados arrojados por este tipo de estudios también deben ser tomados con cautela por varias razones. En primer lugar, una de las desventajas de estas investigaciones fue que las encuestas de los descriptores fueron diseñadas a partir de las tareas que aparecen en el TOEIC. Por ende, dado que los encuestados ya habían presentado

el examen (en un lapso de dos años, en el caso de Powers, 2015, y de unos cuantos días en el estudio de Powers et al., 2010 y Sinharay et al., 2009) y obtenido sus resultados, éstos pudieron haber afectado la autoevaluación, puesto que los examinados ya tenían conocimiento de su desempeño en el examen y, por lo tanto, podían autoevaluarse y decir cómo pensaban que podría ser su desempeño en este tipo de tareas.

De igual forma, otro de los problemas de las autoevaluaciones subyace en la imposibilidad de lograr que todas las personas se evalúen de forma objetiva. Aun cuando en este caso la autoevaluación no conllevaba repercusiones de ningún tipo para los sustentantes, aun así, resulta difícil garantizar que la autoevaluación sea objetiva. De hecho, en el estudio conducido con estudiantes de Latinoamérica (Sinharay et al., 2009), los investigadores encontraron que éstos tendían a sobreestimar sus habilidades comunicativas. Esto sucede con ciertas culturas, mientras que en otras, el fenómeno de autoevaluación puede ser la tendencia a subestimar las habilidades que posee un estudiante Leki et al. (2008). Aunado a esto, también cuando se les pide a los estudiantes que se autoevalúen, en ocasiones hay una distorsión de la realidad cuando éstos saben que sus comentarios pueden incidir en cómo son socialmente aceptados por parte de sus profesores o de los propios investigadores (Powers, 2015).

Finalmente, una de las mayores desventajas de los estudios anteriormente mencionados es que éstos se encuentran fuera de un contexto real de uso de la lengua. En otras palabras, los estudiantes y los docentes encuestados especulan respecto de las tareas comunicativas a las que podrían enfrentarse en el “mundo real”, pero en realidad no se encuentran en un país anglófono desempeñando dichas tareas. Por ende, tanto los resultados como las autoevaluaciones de los examinados y las evaluaciones de los docentes se encuentran aún en un terreno hipotético y distante de la vida real. Entonces, para poder extrapolar eficazmente los resultados de un examen al mundo real, es indispensable que los examinados se encuentren realizando tareas dentro del ámbito para el cual fueron evaluados.

De hecho, Suzuki (2015) menciona que los factores experienciales juegan un papel importante en las autoevaluaciones de los estudiantes. En el estudio que el investigador condujo con estudiantes chinos inmersos en un contexto académico en Japón, encontró que los estudiantes con menores habilidades de lectura tendían a sobreestimar, aunque no de manera sistemática, su desempeño en tareas que no habían realizado, y sobre las cuales

solamente especulaban. Sin embargo, los estudiantes con mayores habilidades y que contaban con una amplia experiencia en un abanico de actividades solían ser más reservados y subestimaban su autoevaluación sobre su desempeño académico, de forma sistemática. El autor menciona que esto sucede debido a que la experiencia realmente proporciona medios para que la autoevaluación sea a conciencia y más precisa, mientras que cuando la autoevaluación es abstracta; es decir, sobre tareas en situaciones imaginarias, ésta resulta menos precisa. Dentro de otras variables que inciden en la autoevaluación, el autor menciona: la ansiedad, factores personales y de actitud, autoestima, la motivación para contestar el instrumento y el dominio de la segunda lengua (Suzuki, 2015, p. 64). Pese a todas estas variables, la autoevaluación en un contexto real resulta un instrumento de utilidad para evidenciar de qué forma correlaciona el puntaje de un ECLAA con el quehacer académico de los estudiantes universitarios.

En esta misma línea, Weigle (2010) condujo un estudio para ver de qué forma otros instrumentos alternos correlacionaban con el resultado obtenido en la prueba discreta del TOEFL iBT. El estudio tuvo lugar con estudiantes no nativohablantes del inglés que cursaban estudios en ocho instituciones distintas de los Estados Unidos. En primera instancia, la autora les administró a los 386 participantes dos tareas de escritura del TOEFL iBT. Posteriormente, se generó una encuesta en línea con cuatro secciones. En la primera, se les solicitó a los estudiantes que autoevaluaran sus habilidades orales, de escritura, lectura y comprensión del inglés. En la segunda sección, se les pedía a los estudiantes que indicaran con qué frecuencia experimentaban ciertos problemas típicos de escritura. Finalmente, en las secciones tres y cuatro de la encuesta se les solicitaba a los estudiantes que indicaran la frecuencia con que experimentaban problemas relacionados con otras habilidades lingüísticas y no lingüísticas (p.ej., manejo del tiempo o entendimiento del tema en cuestión). Además del instrumento de autoevaluación, la autora incluyó la evaluación de docentes y muestras de escritura académica recientes. Para la evaluación docente, se les pidió a los alumnos que proporcionaran el contacto de docentes que estuvieran familiarizados con sus habilidades de escritura. En el caso de las muestras de escritura académica, se les solicitó a los participantes que incluyeran un texto típico de su nivel de dominio y otro que no fuera bueno.

La dinámica que se empleó fue la siguiente: los estudiantes se presentaron a la sesión de las PE, entregaron los ejemplares de sus textos académicos y los datos de contacto de sus profesores, y antes de comenzar la prueba, hicieron la encuesta antes descrita en línea. Los

textos del TOEFL iBT fueron evaluados por dos jueces experimentados, en caso de discrepancia mayor a un punto (aproximadamente 20%), la investigadora reevaluó los textos. Los ejemplares académicos fueron corregidos por evaluadores experimentados mediante una rúbrica que se diseñó especialmente para los fines del estudio. 186 estudiantes contaron con la evaluación de un profesor, 112 con la dos, y 88 sin información.

Las correlaciones entre variables se hicieron mediante el programa estadístico SPSS. Las correlaciones de Pearson entre el puntaje del examen y los instrumentos fueron moderadas. Cabe destacar que las correlaciones más altas se encontraron con mediciones globales sobre el dominio de la lengua más que con aspectos puntuales de habilidades escritas. En resumen, las correlaciones entre el puntaje de la PE del TOEFL iBT y las autoevaluaciones se encontraron en un rango de 0.31 a 0.43 (entre baja y moderada). En cuanto a los problemas de escritura, las correlaciones oscilaron entre 0.17 a 0.29 (entre muy baja y baja). En relación con las evaluaciones de los docentes, las correlaciones se ubicaron entre el 0.13 y 0.46 (entre muy baja y moderada). Finalmente, la comparación con muestras escritas (académicas y los textos del TOEFL iBT) mostró correlaciones bajas entre 0.20 y 0.40.

El estudio anteriormente descrito da cuenta de un enfoque más amplio para poder correlacionar el puntaje obtenido en la PE de un ECLAA y otros indicadores de escritura: evaluaciones de docentes, autoevaluaciones de los estudiantes y ejemplares de textos académicos. Sin embargo, aun así, las correlaciones entre éstos oscilaron entre muy bajas y moderadas. Lo anterior demuestra lo intrincado que ha resultado extrapolar los resultados obtenidos en un ECLAA con lo que un estudiante es capaz de hacer en el “mundo real”. Ahora bien, el estudio de Weigle (2010) presenta algunas limitaciones. Primeramente, los resultados de la PE del TOEFL iBT se compararon con las autoevaluaciones de los estudiantes en las cuatro habilidades. En todo caso, sería mejor restringir la comparación exclusivamente a las habilidades de escritura. De igual forma, el cuestionario de los docentes incluía, además de la habilidad escrita, la habilidad oral de los estudiantes y aspectos generales de dominio lingüístico. El cuestionario docente debería ceñirse exclusivamente a las habilidades de escritura de los estudiantes. Finalmente, en cuanto a la comparación de textos del TOEFL iBT con su contraparte académica, la autora menciona que éstos últimos eran de tan variada naturaleza que ésta pudo resultar en dificultades al momento de evaluarlos. Adicionalmente, no se podía evidenciar el tipo de ayuda que recibieron los estudiantes para producir sus textos, con lo cual no se podía aseverar hasta qué punto el producto final se encontraba influenciado

por el proceso natural de escritura académica (tiempo invertido en la redacción, uso de traductores, diccionarios, correctores electrónicos, y ayuda de terceras personas).

En suma, lo que se puede apreciar de todos los estudios antes mencionados es lo intrincado que resulta extrapolar los resultados de la PE de un ECLAA al “mundo real”. Por una parte, los estudios comparativos entre el puntaje que obtiene un candidato en un ECLAA y sus notas universitarias son epistemológicamente incongruentes, pues el resultado académico es multifactorial y depende en gran medida de la motivación y participación activa de un estudiante. Además, si a esto le añadimos que los estudiantes con puntajes bajos son excluidos del ingreso al ciclo universitario y la atrición del lenguaje desde el momento en que un candidato presenta un examen y el momento en que escribe sus textos académicos es considerable, la comparación conduce a resultados espurios. Por estas razones, los estudios que han buscado extrapolar los resultados de la PE de un ECLAA con otras fuentes provenientes del “mundo real” (p.ej., autoevaluaciones de los sustentantes, evaluaciones de los docentes y cuestionarios que permitan identificar las dificultades de escritura que presentan los estudiantes) resultan más apropiados y congruentes que los mencionados anteriormente. Estos estudios son aún incipientes, pero han abierto nuevos derroteros en las investigaciones que se plantean en relación con la inferencia de extrapolación.

6.3. MÉTODOS E INSTRUMENTOS

En este estudio, se planteó la inferencia relacionada con la extrapolación de los resultados del examen al “mundo real”. Esta inferencia fue la siguiente:

Los resultados de la PE son extrapolables a las habilidades de escritura académica que requieren los estudiantes de posgrado no hispanohablantes

Para poder respaldar esta inferencia mediante evidencia se establecieron las siguientes aseveraciones.

- Los resultados obtenidos por los estudiantes en la PE del EXELEAA correlacionan de forma positiva y significativa con otros instrumentos de evaluación de escritura académica en español.
- Los estudiantes más proficientes demuestran un nivel de escritura académica superior en comparación con aquellos con menor dominio de escritura.

Con base en estas aseveraciones, se diseñó un estudio de tipo mixto con dos propósitos. En primer lugar, se buscó obtener muestras externas al examen que pudieran correlacionarse cuantitativamente con el puntaje obtenido por los estudiantes en la PE del EXELEAA. Para esto, se diseñó una encuesta de autoevaluación sobre las habilidades de escritura académica en español y también una entrevista semiestructurada sobre la escritura académica en español como lengua extranjera. En segundo lugar, se hizo un análisis cualitativo de la entrevista para dar cuenta de la escritura académica en español como LE, su proceso y las variables que influyen al momento de escribir en un contexto académico.

6.3.1. LOS PARTICIPANTES

100 estudiantes universitarios de grado y de posgrado participaron en la PE del EXELEAA; sin embargo, para este estudio se incluyó solamente a los estudiantes de posgrado que participaron en el examen. La razón primordial para elegir exclusivamente a los alumnos de posgrado radicó en que se deseaba contar con estudiantes que ya se encontraran escribiendo en español en el contexto universitario y que pudieran reflexionar sobre su proceso de escritura. Como se mencionó anteriormente, casi la totalidad de los estudiantes extranjeros no hispanohablantes de licenciatura se encuentran de intercambio por un semestre académico; razón por la cual se corría un alto riesgo de que no contaran con experiencia de escritura académica. Esto representaba un requisito indispensable para dar sustento a la inferencia de extrapolación.

Otra de las razones por las cuales se decidió conducir el estudio con estudiantes y no prospectos se debió al fenómeno de atrición. A este respecto, Kokhan (2012, 2013) identificó que, en el caso del inglés, los estudiantes se preparan intensamente antes de presentar un examen de lengua; por lo cual, sus habilidades en la lengua meta se encuentran en el clímax al momento del examen. Sin embargo, tras la prueba, la falta de práctica y el lapso del proceso de admisión –que por lo menos dura dos años– fomentan la pérdida de las habilidades demostradas al momento de haberse sustentado el examen. Esto se debe a que en general los estudiantes presentan el examen en su país natal, y una vez que han conseguido el puntaje deseado, dejan del lado cursos de lengua o de preparación para un examen (p.ej., TOEFL o IELTS) y no hacen uso de la lengua meta hasta que ingresan a la universidad. Por ende, cuando un estudiante comienza a escribir en el contexto universitario, sus habilidades no son las mismas a los resultados que originalmente había arrojado el examen de lengua.

El proceso de identificación y selección fue el siguiente. En agosto de 2013 y 2014, se le solicitó a la Unidad de Posgrado su participación en el estudio. Mediante un oficio, se le pidió al posgrado la base de datos de todos los estudiantes de posgrado no hispanohablantes que se encontraban registrados. Esta base de datos incluía tanto a los estudiantes de nuevo ingreso como a los de reingreso. En la base datos se incluyó la siguiente información: nombre, nacionalidad, edad, nivel de estudios cursado, programa de posgrado inscrito, correo electrónico y año de ingreso a la UNAM. Una vez que se obtuvo esta información, se les envió un correo electrónico a los estudiantes con una invitación para que participaran en el pilotaje (administrado a finales de noviembre de 2013) y en la primera aplicación del examen (administrada a principios de diciembre de 2014). En la invitación se especificó una fecha y horario determinado para el examen, así como también los beneficios a los que serían acreedores por participar en el examen.

De hecho, probablemente la gran mayoría de los estudiantes convocados ya contaban con un certificado oficial de español como lengua extranjera (expedido por el CEPE), pues los posgrados generalmente lo solicitan como requisito de ingreso a la universidad. Por ello, con el fin de incentivar la participación de los estudiantes en el examen y en el estudio cualitativo y cuantitativo, al primer grupo de estudiantes de posgrado que participó en el pilotaje del examen, a finales de noviembre de 2013, se le ofreció un taller gratuito de escritura académica, posterior al examen, durante una semana. La invitación a la encuesta se les envió al término del examen y las entrevistas se llevaron a cabo durante la semana del taller que se ofreció. Sin embargo, de los 20 estudiantes que participaron en el examen, solamente siete asistieron al taller, cinco participaron en las entrevistas y ocho en el cuestionario.

Al año siguiente, cuando se convocó nuevamente a los estudiantes a participar en el estudio, como incentivo, se les ofreció el examen de español (de carácter oficial) de forma gratuita a cambio de contestaran una encuesta en línea sobre sus habilidades de escritura académica y también a que participaran en una entrevista sobre su proceso de escritura académica y variables asociadas a ésta. La invitación a contestar la encuesta se les envió una semana previa al examen y como recordatorio, un día después del examen. Las entrevistas se llevaron a cabo el mismo día del examen (antes o después), esto con el fin de captar al mayor número de estudiantes. De los 25 estudiantes que sustentaron el examen, 20 participaron en la entrevista y 17 en el cuestionario. Entonces, de un total de 45 estudiantes de posgrado no

hispanohablantes, 22 contestaron la encuesta en línea y 25 participaron en la entrevista semiestructurada (ver Anexo 3.1).

6.3.2. PRIMERA PARTE: LA CORRELACIÓN DE LA PE DEL EXELEAA CON OTROS INSTRUMENTOS DE EVALUACIÓN DE ESCRITURA ACADÉMICA

El puntaje de los estudiantes obtenido en la PE del EXELEAA (ver Capítulo 5, Estudio 2) fue correlacionado con el resultado obtenido en la encuesta de autoevaluación y en la evaluación de la entrevista semiestructurada sobre el nivel de escritura académica de este presente estudio.

6.3.2.1. LA ENCUESTA DE AUTOEVALUACIÓN DE LAS HABILIDADES DE ESCRITURA ACADÉMICA

El estudio cuantitativo consistió en que los estudiantes identificaran el nivel de competencia que consideraban tener en las habilidades de escritura académica consideradas como importantes para el éxito académico (ver Anexo 3.2). En la primera sección se les pidieron datos personales, y en la encuesta, se incluyeron los mismos 33 descriptores de escritura académica que les fueron administrados a los docentes del posgrado (ver Anexo 1.3), pero en vez de que se evaluara la importancia de una habilidad, a los estudiantes se les pedía que identificaran el dominio que consideraban tener en cada una de estas habilidades de escritura académica. Entonces, la convención que se empleó fue la siguiente: 0 (no se cuenta con la habilidad de escritura requerida), 1 (nivel bajo), 2 (nivel suficiente), 3 (nivel intermedio), 4 (nivel alto), 5 (nivel avanzado).

6.3.2.2. LA ENTREVISTA SEMIESTRUCTURADA

Para el estudio cualitativo, se diseñó una entrevista semiestructurada partiendo de ciertas variables que se suponía se encontraban asociadas al dominio de la escritura académica. La razón para optar por una entrevista semiestructurada se fundamentó en la necesidad de conocer el dominio de las habilidades de escritura de los estudiantes no hispanohablantes del posgrado, así como su percepción de la escritura y de las variables asociadas a ésta. Con base en un primer guion de la entrevista, en diciembre de 2013 se entrevistó a cinco estudiantes de posgrado que participaron en el pilotaje del EXELEAA. Con esta primera muestra, se hizo una transcripción en Word y una tabla en Excel para poder llevar a cabo un análisis preliminar de las categorías y unidades de análisis. Un grupo de cuatro colegas y una especialista en análisis de entrevistas revisaron la categorización y, a partir de ahí, se afinó el guion de la entrevista

(ver Anexo 3.3), con el cual fueron entrevistados los 20 estudiantes que participaron en 2014. El investigador entrevistó y grabó a los estudiantes en formato digital sin compresión de datos (WAVE) y bajo su previo consentimiento.

Durante la entrevista, dependiendo de las respuestas del entrevistado, se elaboraron otras preguntas que resultaban de interés para la investigación. Esto debido a que a veces se buscó profundizar en un tema, mientras que en otras, se abrieron nuevos panoramas que resultaron interesantes de explorar. También el orden de las preguntas se mantuvo flexible, puesto que en ocasiones los estudiantes respondían de forma simultánea otras preguntas posteriores sin que se formulara la pregunta. En estos casos, dependiendo del tema que abordaban, se alteró el orden de las preguntas. Finalmente, se consideraron algunos conceptos que Bryman (2012, p. 475) contempla para una entrevista exitosa; a saber: ser conocedor del tema, tener estructura y claridad, ser gentil, sensible, abierto y crítico, y saber recordar e interpretar la información. En este sentido, y retomando lo que sugiere el autor y tomando en consideración que se trataba de estudiantes no hispanohablantes, se trató de evitar que las preguntas fueran rebuscadas o que se usaran términos que resultaran poco familiares para los entrevistados. También se ofreció ayuda cuando tenían dificultades en encontrar las palabras que buscaban para expresarse.

6.3.2.3. LA RÚBRICA ANALÍTICA PARA EVALUAR LA ENTREVISTA

Con el fin de poder transformar la información recolectada de las entrevistas en datos cuantitativos, se diseñó una rúbrica analítica para evaluarlas. El objetivo de este proceso fue que se pudieran correlacionar los resultados obtenidos por los estudiantes en la PE del EXELEAA con los de la entrevista. La rúbrica fue diseñada por el investigador y revisada por la responsable del examen. En la escala se incluyeron 12 categorías y cuatro niveles de dominio (0 al 3). Dado que no todos los estudiantes respondieron a cada una de las categorías incluidas en la rúbrica, se decidió añadir un rubro más (denominado: no aplica) cuando no se contaba con información para evaluar su proceso de escritura. Con el fin de evaluar su funcionamiento, la responsable del examen y el investigador evaluaron, de forma independiente, cinco entrevistas previamente seleccionadas al azar. Tras este pilotaje, se revisaron los resultados de las cinco entrevistas, pero no se consideró hacer ninguna modificación a la escala de evaluación (ver Anexo 3.4). Posteriormente, se prosiguió con la evaluación de los otros 20 estudiantes, también de forma independiente.

En caso de discrepancia mayor al 10 por ciento, intervino un tercer evaluador; es decir, se contabilizó el resultado otorgado por cada uno de los evaluadores y se obtuvo la media de cada examinado. Entonces, la media de cada examinado podía oscilar entre 0 y 3. El 10 por ciento equivale a ± 0.3 , por lo cual cuando el resultado entre evaluadores excedía dicho valor, se decidió que interviniera un tercer evaluador. El tercer evaluador era una especialista en evaluación cualitativa y no tenía conocimiento alguno de la investigación, razón por la cual se le tuvo que dar una sesión de inducción al uso de la rúbrica y a la evaluación de las entrevistas. De las 25 entrevistas evaluadas, la evaluadora tuvo que evaluar 8 en las cuales se presentó más de un 10 por ciento de discrepancia. En estos casos, el puntaje final del examinado se obtuvo mediante el promedio de los tres evaluadores.

6.3.2.4. EL ANÁLISIS CUANTITATIVO MEDIANTE EL SOFTWARE SPSS

Una vez que se recolectaron los datos de la encuesta y de la evaluación de la entrevista, se obtuvieron las puntuaciones medias de cada uno de los estudiantes en ambos instrumentos. Posteriormente, se generó una tabla, en SPSS versión 21, para correlacionarlos y poder dar cuenta de cómo se relacionan los puntajes de la PE del EXELEAA con los datos de la encuesta y de la entrevista (instrumentos ajenos al examen y relacionados con el uso de la lengua en un contexto real).

6.3.3. SEGUNDA PARTE: LA ESCRITURA ACADÉMICA DE LOS ESTUDIANTES NO HISPANOHABLANTES

En esta parte del estudio se hizo un análisis cualitativo de la entrevista semiestructurada de acuerdo a cómo se clasificó a los estudiantes, en grupos, dependiendo de su nivel de escritura: intermedio bajo, intermedio alto, avanzado.

6.3.3.1. LA ENTREVISTA SEMIESTRUCTURADA

La gran ventaja de las entrevistas, en general, es que ayudan a recolectar información relacionada con las experiencias de las personas, sus percepciones, actitudes o sentimientos con respecto a cierta temática. Zhang y Wildemuth (2009) explican que existen tres tipos de entrevistas: estructuradas, semiestructuradas y no estructuradas. Dependiendo del objetivo de la investigación, se emplea un tipo u otro. Por ejemplo, las entrevistas estructuradas funcionan como encuestas en las que se espera una respuesta abierta en vez de que el entrevistado seleccione entre varias opciones. Las entrevistas estructuradas contienen un guion, el cual se sigue al pie de la letra, que no es flexible ni da posibilidad a que se puedan preguntar temas

adicionales que podrían resultar de interés. Debido a su rigidez, éstas permiten que cualquier persona con entrenamiento mínimo pueda aplicarlas, pues prácticamente se trata de un cuestionario o encuesta oral y con respuestas cerradas (p.ej., de opción múltiple) o abiertas, pero cortas. En el caso de las entrevistas semiestructuradas, pese a que se cuenta con un guion previamente establecido, éste es flexible y permite que el investigador profundice más en un aspecto relevante o interesante para su investigación. De igual manera, es posible que nuevos temas o conceptos emerjan durante las entrevistas. Naturalmente, debido a que se cuenta con un guion flexible, no es tan fácil que sea operada por personas que se encuentran fuera del proceso de investigación. Finalmente, las entrevistas no estructuradas son aquellas en las cuales se pretende “capturar una realidad social”. También se la conoce como entrevista conversacional informal, entrevista no estandarizada o entrevista etnográfica (Zhang y Wildemuth, p. 222). En este tipo de entrevistas, ni la pregunta ni las categorías se encuentran predeterminadas. Debido a que en la presente investigación se contaba ya con ciertos aspectos relacionados al proceso de escritura de un no nativohablante, así como un guion definido, se optó por una entrevista semiestructurada.

6.3.3.2. EL PROCESO DE TRANSCRIPCIÓN DE LAS ENTREVISTAS

Al momento de concebir las entrevistas, también es importante tener en mente el proceso de transcripción, pues éste generalmente conlleva una cantidad de tiempo considerable. Sin embargo, esto no exige que las transcripciones o porciones de éstas sean lo más detalladas que se pueda, incluso palabra por palabra (Knobel y Lankshear, 2001). De hecho, estos autores agregan que incluso las pausas, el alargamiento de vocales dentro de una palabra y las interrupciones deberían ser codificadas en una transcripción. En este sentido, Bryman (2012) hace énfasis en que no sólo nos interesa qué se dice, sino también cómo se dice; por eso, es importante contar con convenciones que nos ayuden a poder rescatar la manera en la cual se dicen las cosas. Por el contrario, dependiendo del tipo de estudio del cual se trate, Zhang y Wildemuth (2009) mencionan que se debe ponderar la necesidad de hacer o no una transcripción cabal de las entrevistas, pues hay casos en los que la información es demasiada o no siempre es relevante para que se haga una transcripción completa y literal. En el caso del presente estudio, tomando en consideración lo que los autores antes mencionados sugieren, se hizo una transcripción sumamente detallada, aunque no cien por ciento palabra por palabra.

Por otra parte, Bazely y Jackson (2013) recomienda que el investigador sea quien haga la transcripción de las entrevistas o por lo menos la revisión de éstas, puesto que de esa manera

podrá captar mejor en el texto aspectos no lingüísticos expresados por los entrevistados (p.ej., entonación, ironía, sarcasmo, etc.). Debido a que lo anterior no fue posible, poco más de la mitad de las entrevistas fueron transcritas por dos estudiantes de sociología a quienes se les pagó por contrato de obra terminada. Una vez transcritas, fueron revisadas y editadas por un asistente de servicio social del CEPE y el investigador. Se buscó que las transcripciones capturaran la mayor cantidad de información posible, aunque éstas no se hicieron de forma literal. En caso de errores gramaticales o de pronunciación dichos por los entrevistados, se decidió conservarlos y colocar entre corchetes lo que se había querido decir. Entre paréntesis aparece información no verbal capturada al momento de la transcripción.

6.3.3.3. EL ANÁLISIS DE LOS DATOS

Vale la pena destacar que los estudios cualitativos han sido duramente criticados por la supuesta falta de rigor metodológico para diseñarlos, implementarlos y analizarlos. Sin embargo, es importante aclarar que el fin de éstos y los cuantitativos es diferente, y más bien resultan complementarios para entender una problemática dada. De hecho, Creswell y Plano Clark (2011) insisten en la utilidad de llevar a cabo estudios mixtos, pues afirman que éstos nos ayudan a comprender fenómenos desde distintas aristas. Al igual que en el análisis cuantitativo, los principios de transferencia también aplican para el análisis cualitativo; es decir, los métodos utilizados deberán poder emplearse en otros contextos similares. Para ello, primeramente, se debe demostrar que los estudios son plausibles no solamente en la manera en la cual se fundamentan, sino también en cómo se realizan. Por esta razón, en los estudios cualitativos, al igual que en los cuantitativos, es importante considerar la noción de validez. Knobel y Lankshear (2001) y Zhang y Wildemuth (2009) mencionan que los conceptos de validez (tanto interna como externa) y de confiabilidad también aplican a la sistematización en el diseño, recolección, análisis e interpretación de los datos; así que, los autores hacen hincapié en la necesidad de tener claridad sobre la pregunta de investigación antes de recolectar los datos de la investigación.

En los estudios de corte cualitativo, dependiendo del tipo de razonamiento –deductivo o inductivo⁵–, se determina el enfoque y la metodología que se empleará. Thomas (2006)

⁵ Thomas (2006, p. 238) define “el enfoque inductivo como un procedimiento sistemático para analizar datos cualitativos en los cuales el análisis se encuentra guiado por los objetivos específicos de la investigación”. En el caso del análisis deductivo, Thomas (2006, p. 238) agrega que “se refiere al análisis de datos que se establece para probar si éstos son consistentes con suposiciones previas, teorías o hipótesis identificadas o construidas por el investigador”.

describe cuatro enfoques para poder abordar estudios cualitativos: enfoque general de tipo inductivo, teoría fundamentada, análisis del discurso y estudios fenomenológicos. En un análisis más fino, Zhang y Wildemuth (2009, p. 309) describen tres enfoques distintos de análisis inductivo. El primero es el análisis del contenido, éste es el tipo de análisis tradicional en el cual las categorías emergen directamente y de forma inductiva a través de los datos. En este proceso se hace una comparación constante de datos mediante un cuidadoso escrutinio de éstos. El segundo se conoce como análisis de contenido dirigido y se basa previamente en una teoría para poder efectuar la codificación. El fin de este enfoque es poder validar o extender teorías. El tercer enfoque es el análisis de contenido sumativo, en el cual se inicia aparentemente de forma cuantitativa, pues se empieza con el conteo de palabras; sin embargo, el fin es explorar la frecuencia de aparición de éstas de manera inductiva. Debido a que en la presente investigación no se contaba con una teoría o modelo previo de escritura de segundas lenguas que se pretendiera validar, se adoptó el enfoque general de tipo inductivo mediante un análisis tradicional de datos.

6.3.3.4. EL ANÁLISIS DEL CONTENIDO

El análisis del contenido resulta primordial para conferirle validez a los estudios cualitativos. Para poder realizar eficazmente dicho análisis –cuando se trata de entrevistas semiestructuradas–, es importante que se cuente con preguntas de investigación acotadas que ayuden a clasificar la información respecto de lo que se desea investigar. Andréu (2001) y Zhang y Wildemuth (2009) comentan que es importante definir primeramente la unidad de análisis, de la cual se desprenden 3 subcategorías: unidad de muestreo, la unidad de registro y la unidad de contexto. La unidad de muestreo hace referencia al conjunto de datos orales o escritos a partir de los cuales se lleva a cabo el análisis. En la presente investigación, esta unidad de muestreo hace referencia a las entrevistas con estudiantes de posgrado no hispanohablantes que participaron en el EXELEAA. La unidad de registro corresponde a las porciones del discurso que son susceptibles a categorizarse, según los fines de la investigación. Esta parte es compleja y no siempre es fácil decidir cuál será la unidad de registro, pues en ocasiones puede ser una palabra, una frase o un párrafo. Lo más importante es distinguir entre una unidad lingüística (p.ej., una palabra, oración o párrafo) de una unidad temática. En el segundo caso, se prioriza la expresión de una idea que puede estar contenida en una palabra, en un segmento de oración o de párrafo. Para el presente estudio, los codificadores acordaron que la unidad de registro sería temática; es decir, se concibió el

contenido mínimo que diera información completa a la pregunta formulada u otros conceptos que emergieran de la entrevista. Finalmente, la unidad de registro también se encuentra inserta en la unidad del contexto; por eso, muchas veces la selección de una sola palabra o frase no es suficiente para que la unidad de registro sea comprendida en su totalidad. De ahí que sea común que muchas veces se necesite contextualizar la unidad de registro, con el fin de poder hacer una correcta interpretación del contenido.

6.3.3.5. EL PROCESO DE CATEGORIZACIÓN

Para poder realizar el análisis de los datos, éste se hace mediante la previa identificación de categorías⁶. Rose y Sullivan (1996, citados en Knobel y Lanshear, 2001) afirman que “una categoría es un agrupamiento de datos que son similares, parecidos u homogéneos” (p. 82). El análisis categórico puede llevarse a cabo de forma inductiva o deductiva, dependiendo de si la investigación se concibe a partir de una teoría previa o no. En el análisis inductivo, es común que la constante relectura de transcripciones dé pie a la elucubración de temas y categorías de análisis, con los cuales se pueden establecer los marcos de codificación de la información. Las categorías también guardan una relación entre sí a partir de las cuales es posible configurar una red conceptual que ayuda a comprender la interacción que surge entre ellas. Dependiendo de la relación que guardan las categorías, éstas pueden agruparse bajo otras denominadas superordinadas. La ventaja principal de contar con estas últimas es que permiten agrupar categorías y así hacer los datos más asequibles y fáciles de interpretar (Knobel y Lanshear, 2001, p. 89).

Thomas (2006, p. 40) menciona que las categorías que resultan de la codificación de datos generalmente poseen las siguientes características:

- La etiqueta de la categoría (una palabra o frase corta se usa para referirse a una categoría).
- La descripción de la categoría.
- Texto asociado a la categoría.
- Enlaces que establecen relaciones entre distintas categorías (p.ej., diagramas en donde se percibe la jerarquía entre categorías).

⁶ Bazeley (2009, p. 6) hace una distinción entre temas, categorías y conceptos, pues es común que algunas de éstas se utilicen de forma intercambiable. Las categorías hacen referencia al proceso descriptivo de codificación, mientras que los conceptos se encuentran en un plano abstracto y no codificable. El tema generalmente se usa para describir la relación que se da entre los datos.

- El modelo en el cual se inserta la categoría (i.e., el sistema de categorización puede emplearse posteriormente para ser incorporado en un modelo, teoría o marco).

6.3.3.6. EL PROCESO DE CODIFICACIÓN

Para poder analizar la información, primero es importante codificarla. Este proceso comienza con una minuciosa lectura de los datos con el fin de extraer unidades significativas que permitan la creación de distintas categorías. La unidad de análisis, cualquiera que ésta sea (p.ej., palabra, frase, oración, párrafo), se asigna a las distintas categorías previamente establecidas. Andréu (2001) hace referencia a la codificación del material de análisis y menciona que “la codificación consiste en una transformación mediante reglas precisas de los datos brutos del texto” (p. 14). Al realizar esta decodificación, los datos se convierten en unidades significativas que pueden ser interpretadas de acuerdo con los fines de la investigación. De acuerdo con Bardin (1996, citado en Adréu, 2001), en la codificación es importante tomar en consideración las siguientes reglas para el recuento de las unidades de análisis: la presencia, la frecuencia, la intensidad, la dirección, el orden y la contingencia de los datos a analizar.

Cuando se lleva a cabo el proceso de codificación de los datos, es frecuente que éstos quepan en diferentes categorías. Consecuentemente, las categorías no siempre son mutuamente excluyentes y es común que éstas en ocasiones se traslapen. Zhang y Wildemuth (2009) comentan que esto es normal en estudios cualitativos, pues a diferencia de los estudios cuantitativos, los datos pueden pertenecer a distintas categorías. Dependiendo de qué tanto se traslapen las categorías, puede ser que algunas se puedan suprimir, redefinir o incluso agregar nuevas. Sin embargo, es importante que de forma previa al proceso de categorización se tenga en cuenta de forma clara cuál es el fin de la investigación.

En el proceso de codificación, Thomas (2006), menciona los siguientes pasos:

- Preparación o limpieza de datos. Básicamente se refiere al proceso de transcripción de cada una de las entrevistas.
- Lectura detallada del texto.
- Creación de categorías.
- Traslape de texto. Las categorías no siempre son mutuamente excluyentes; por lo que es probable que en algunos casos la información se traslape entre categorías.

- Texto no codificado. Es muy probable que una gran cantidad del texto no sea asignado a ninguna categoría, pues simplemente no todo es relevante para los objetivos de la investigación.
- Revisión continua y refinamiento de las categorías. Durante el análisis, es común que algunas categorías se dividan o se fusionen. También durante esta fase se comienzan a establecer los vínculos y relaciones entre distintas categorías.

6.3.3.7. EL PROCESO DE INTERPRETACIÓN

El análisis cualitativo no solamente implica hacer una descripción de los datos; de hecho, Bazeley (2009) apunta que la descripción por sí sola no es suficiente. Los datos deben trabajarse de tal forma que se pueda establecer un vínculo entre ellos con el fin de que éstos aporten un valor informativo completo. Por ello, es importante describir los datos, pero también compararlos y relacionarlos entre sí para poder comprender un fenómeno o explicar un modelo o teoría. Por tal razón, lo primero que se hizo fue hacer una división de grupos. Es decir, con base en la rúbrica que se empleó para evaluar las entrevistas, se determinaron cuatro grupos: bajo, intermedio bajo, intermedio alto y avanzado. Debido a que ningún estudiante cayó dentro del grupo bajo, el resto fue clasificado en los tres grupos restantes. La razón fundamental para efectuar tal división estribó en que en el diseño inicial de la investigación se especuló que el reporte de cada uno de estos grupos debería presentar patrones similares en cuanto al dominio de las habilidades de escritura y a las variables asociadas a éste. Una vez que se revisaron los datos en cada uno de estos grupos, se detectaron aquellas constantes dentro del proceso de escritura y también las variables asociadas al proceso de ésta. Por último, se compararon los resultados entre grupos para poder llevar a cabo interpretaciones sobre los procesos de escritura dependiendo del nivel de dominio de la lengua meta.

6.3.3.8. EL SOFTWARE NVIVO PARA EL ANÁLISIS CUALITATIVO

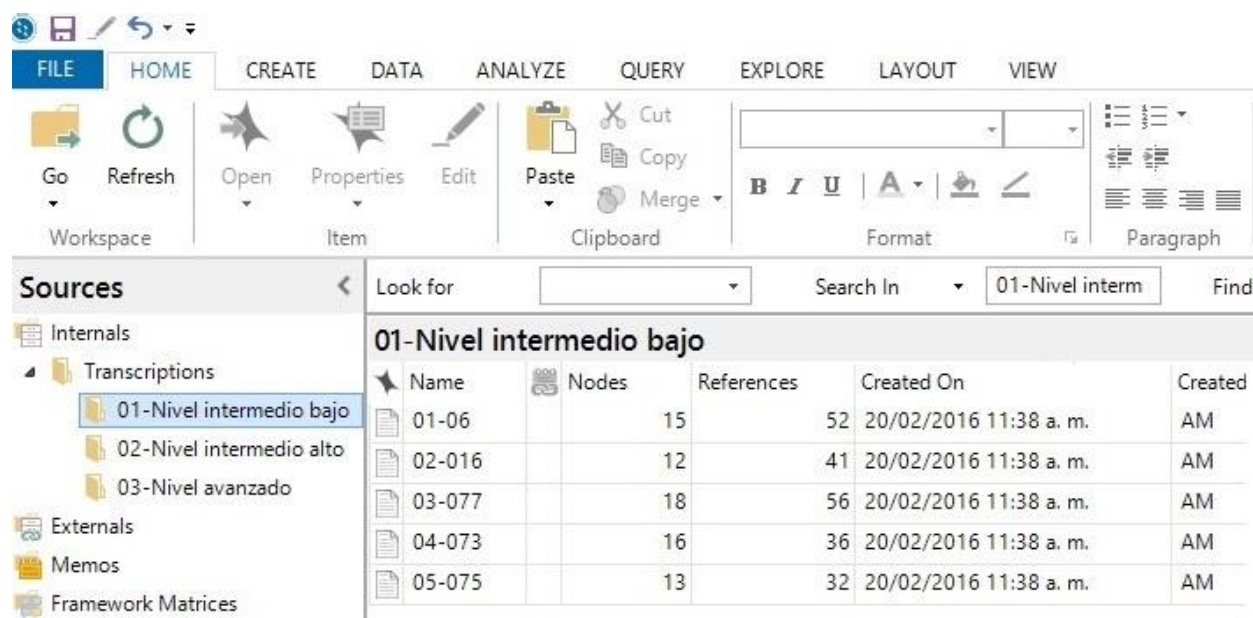
El análisis cualitativo de las entrevistas se hizo mediante el software NVivo versión 11. Debido a lo importante que hoy en día es gestionar grandes cantidades de información proveniente de estudios cualitativos, resulta importante considerar el uso de programas computacionales que no solamente nos permitan gestionar la información de forma eficiente, sino que además permitan codificar la información de forma ordenada, elaborar mapas mentales, identificar la frecuencia de palabras y establecer las correlaciones entre codificadores (véase Zhang y Wildemuth, 2009). Esto no solamente facilita el proceso de análisis, sino que además su uso es importante para lograr la confiabilidad. Adicionalmente, una de las grandes

ventajas de NVivo es que nos permite construir matrices para poder observar la coocurrencia de códigos o unidades de análisis dentro del texto (Bazeley, 2009).

Bazeley y Jackson (2013) y Bernauer, Lichtman, Jacobs y Robertson (2013) arguyen que este programa constituye una herramienta de apoyo para el pensamiento crítico. Agregan que es importante reconocer que en el centro de la investigación se encuentra el investigador que hace uso del programa, y que el programa por sí mismo no proporciona ningún fundamento metodológico para conducir la investigación, ni se encuentra dirigido a ningún tipo de investigación cualitativa en particular. En otras palabras, si el investigador no cuenta con el conocimiento suficiente de lo que desea investigar, el programa será de poca ayuda, como sucede con cualquier otra herramienta tecnológica. Los programas computacionales no son creativos ni analíticos por sí mismos, por lo cual requieren ser manipulados eficientemente para que los resultados obtenidos sean congruentes con las expectativas e interrogantes previamente planteadas al inicio de la investigación. A esto, Creswell (2013) añade que es el investigador quien hace la codificación, no el programa. De igual forma es él quien establece las categorías y las posibles relaciones entre categorías.

6.3.3.9. DESCRIPCIÓN DEL FUNCIONAMIENTO DEL PROGRAMA NVIVO

A continuación, se hará una descripción de cómo funciona el programa. Para ello, se tomará el estudio que se llevó a cabo con las entrevistas de los estudiantes del posgrado. NVivo funciona con texto, audios, imágenes y videos, aunque para poder hacer el proceso de codificación directamente, sólo se trabajó con las transcripciones. Para comenzar, se deben copiar las transcripciones en NVivo. En este caso, dado que se trataba del análisis de tres grupos distintos: intermedio bajo, intermedio alto y avanzado, se organizaron las transcripciones en tres subcarpetas con los grupos antes mencionados. La Figura 6.1 muestra del lado izquierdo la carpeta de las transcripciones con sus respectivas subcarpetas y estudiantes (nivel intermedio bajo, nivel intermedio alto y nivel avanzado). En el centro de la pantalla se muestra el archivo (en número) de cada uno de los estudiantes, en este caso del grupo intermedio bajo. También se muestra cuándo se creó el archivo y las iniciales de quien lo creó.

Figura 6.1: Organización de las transcripciones en carpetas

Una vez que se cuenta con las transcripciones dentro de NVivo, es el momento de identificar las categorías de análisis. En una fase inicial, se trabajó con cinco entrevistas pertenecientes a la primera muestra de estudiantes que sustentó el EXELEAA en 2013. Con la información recabada en estas entrevistas, se identificaron las categorías de análisis preliminares y se rediseñó el guion de la entrevista.

Posteriormente, en una segunda fase, el investigador hizo una lectura minuciosa de las 25 entrevistas con el fin de determinar la lista final de categorías. Paralelamente, la especialista en evaluación cualitativa se dio a la tarea de hacer una lectura detallada de las entrevistas con el fin de comprender la naturaleza de la investigación. Una vez concluido este proceso, se llevó a cabo una sesión de inducción para la especialista, quien no estaba familiarizada con el tema de estudio. En esta sesión se discutieron las categorías previamente concebidas por el investigador, pero se dejó abierta la posibilidad de modificarlas. Posteriormente, cuatro entrevistas fueron codificadas en NVivo de forma independiente por el investigador y la especialista con el fin de ver si las categorías previamente definidas eran suficientes, claras y necesarias. Con base en esta experiencia, se redefinieron las categorías y se evaluaron las 25 entrevistas. Durante el proceso de codificación, es probable que emerjan nuevos códigos o se fusionen otros ya existentes. Este caso no fue la excepción, así que fue importante revisar los

nuevos conceptos y categorías que fueron emergiendo de los datos, con lo cual se definieron las siguientes categorías de análisis:

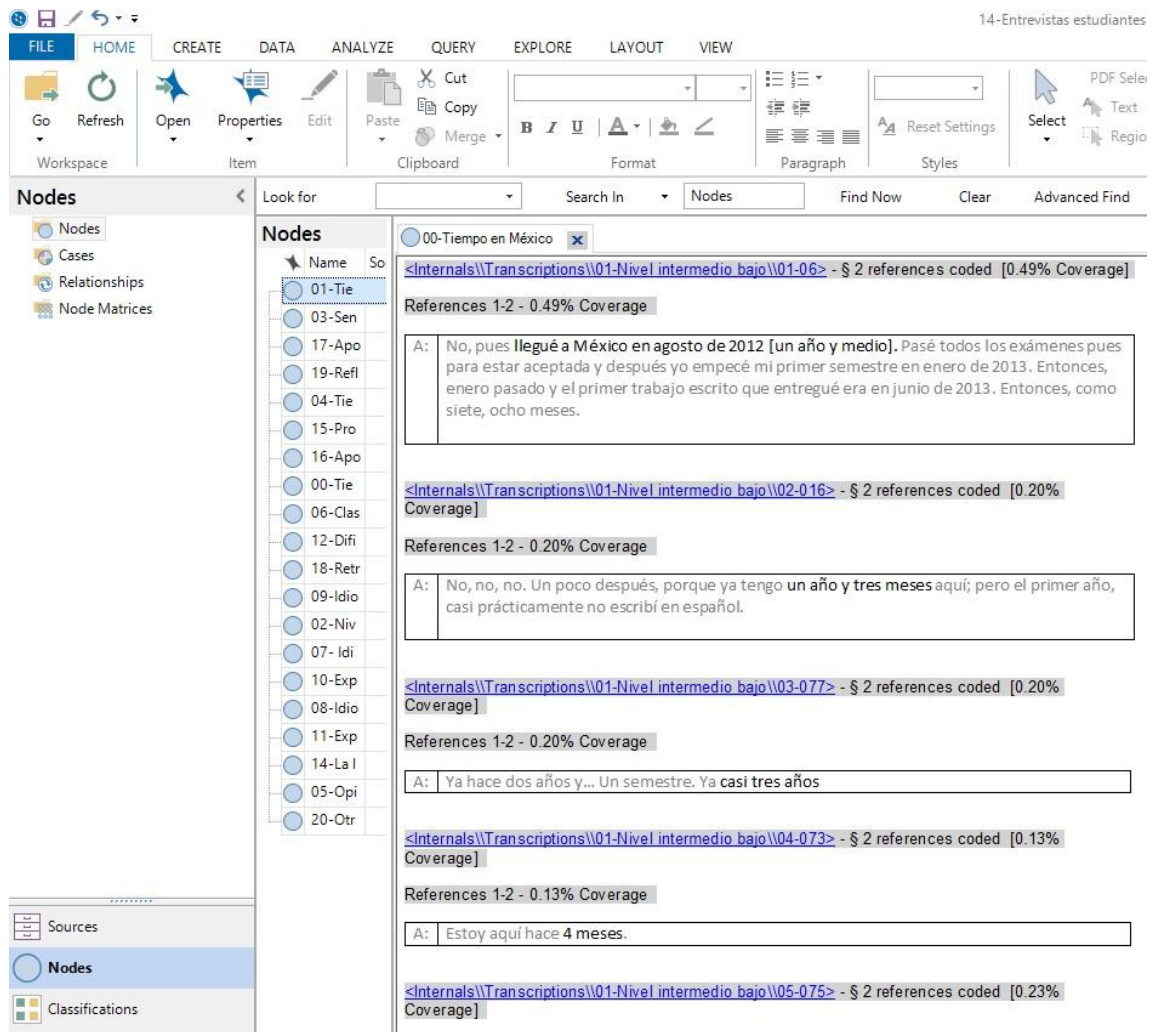
1. Tiempo en México
2. Tiempo de escribir en un contexto académico en español
3. Autopercepción del nivel de escritura en español
4. Sensaciones al momento de escribir en español
5. Tiempo que le toma escribir en español
6. Toma clases de lengua o manifiesta una opinión al respecto
7. Clases que ha tomado o tomó en un contexto académico en español
8. Idioma de las lecturas asignadas en las clases
9. Idioma de las lecturas optativas o por cuestiones de la investigación
10. Idioma de notas de clase
11. Experiencia con exámenes académicos escritos
12. Experiencia con trabajos académicos escritos
13. Dificultades al escribir en español
14. Influencia de la lengua materna para escribir en español
15. Descripción del proceso de escritura
16. Uso de recursos materiales
17. Edición con ayuda de terceros
18. Retroalimentación que recibe sobre su escritura en español
19. Reflexión sobre qué podría mejorar su escritura académica
20. Otros conceptos no incluidos

Tras el análisis de los datos, se decidió fusionar las categorías 15, 16 y 17, pues las tres se encontraban relacionadas dentro de la supercategoría: proceso de escritura. En NVivo, las categorías se denominan nodos. Estos nodos son independientes entre sí, aunque también se pueden agrupar y jerarquizar dentro de otros. A esta relación entre jerarquías se le conoce como árbol. La Figura 6.2 muestra las 20 categorías, en Nvivo, que se identificaron para el Estudio 3. Como se puede apreciar en el centro de la figura, en cada uno de los nodos aparece cuántos entrevistados fueron catalogados en dicho nodo. En el caso de las referencias, se contabiliza a ambos codificadores. También se puede identificar quién creó el nodo y la fecha de creación.

Figura 6.2: Categorías o nodos identificados para las entrevistas

Name	Sources	References	Created On	Created
00-Tiempo en México		23	21/02/2016 11:44 a. m.	AM
01-Tiempo de escritura en un contexto académico en español		25	20/02/2016 11:40 a. m.	AM
02-Nivel de español		20	20/02/2016 11:41 a. m.	AM
03-Sensaciones al momento de escribir en español		25	03/03/2016 10:52 p. m.	AM
04-Tiempo que le toma escribir en español		23	20/02/2016 11:42 a. m.	AM
05-Opinión sobre cursos de lengua		7	20/02/2016 11:44 a. m.	AM
06-Clases que toma en el posgrado		22	20/02/2016 11:42 a. m.	AM
07- Idioma de las lecturas asignadas en clase		19	20/02/2016 11:44 a. m.	AM
08-Idioma de lecturas académicas preferido o de material de investigación		17	20/02/2016 11:45 a. m.	AM
09-Idioma de notas de clase		21	20/02/2016 11:45 a. m.	AM
10-Experiencia y dificultades con exámenes académicos escritos		19	20/02/2016 11:47 a. m.	AM
11-Experiencia y dificultades con trabajos académicos		17	20/02/2016 11:46 a. m.	AM
12-Dificultades al escribir en español		22	21/02/2016 12:15 p. m.	AM
14-La lengua materna ayuda		15	28/02/2016 02:31 p. m.	AM
15-Proceso de escritura		23	20/02/2016 11:47 a. m.	AM
16-Apoyo de recursos materiales		23	20/02/2016 11:48 a. m.	AM
17-Apoyo de amigos, tutores, pareja o lectores		24	20/02/2016 11:48 a. m.	AM
18-Retroalimentación de la lengua		22	21/02/2016 12:21 p. m.	AM
19-Reflexión sobre mejora de escritura académica		24	20/02/2016 11:49 a. m.	AM
20-Otros conceptos no incluidos		6	20/02/2016 11:50 a. m.	AM

Una vez que se contaba con las transcripciones y los nodos, el investigador y la especialista en evaluación cualitativa se dieron a la tarea de codificar cada una de las 25 entrevistas. En NVivo, la codificación se lleva a cabo seleccionando la información de la transcripción y arrastrándola al nodo en donde se desea ubicar. Esto permite copiar la información seleccionada de la transcripción en el nodo deseado. Dado que copia la información, ésta puede codificarse posteriormente en otro nodo si así se requiere. Una ventaja de NVivo es que además de seleccionar la información que el codificador elige, también se selecciona todo el párrafo en el cual se encuentra la información codificada. Esto permite contar con la referencia contextual de donde se obtuvo la información, lo cual facilita el análisis e interpretación de los datos. La Figura 6.3 presenta la primera categoría: tiempo en México. En ésta se puede apreciar como la información de cada entrevistado se encuentra ordenada de acuerdo con su nivel de dominio; es decir, primero los estudiantes con menor dominio de escritura. También se puede ver cómo la información seleccionada se encuentra en negro, mientras que el párrafo del cual fue seleccionada, en gris.

Figura 6.3: Selección de la información dentro de una categoría o nodo

Dado que dos o más personas podrían estar codificando la información, NVivo permite contrastar la información codificada por dos o más personas: consistencia intercodificador. Este contraste se manifiesta en el coeficiente de correlación (Kappa de Cohen). La correlación muestra qué porcentaje del texto codificado coincide entre los dos o más codificadores. Es decir, si dos codificadores seleccionan la misma información de un entrevistado en un nodo, tendrán una correlación positiva de 1. Si ambos codificadores deciden que no hay información que codificar de una entrevista en una categoría, también se asigna la correlación de 1. Si la codificación de la información solo es parcialmente similar entre ambos codificadores, entonces la correlación oscila en un rango de 0.01 a 0.99, dependiendo de qué porcentaje de la selección es igual. Si la información de la entrevista es codificada en una categoría sólo por uno de los codificadores, entonces la correlación es de 0. Finalmente, si los codificadores

seleccionan porciones diferentes de la información en una categoría, entonces la correlación es negativa en un rango de - 0.01 a - 1.

En el caso de la presente investigación, estas correlaciones fueron útiles para poder llevar a cabo sesiones de retroalimentación en las cuales se discutieron las posibles causas de discrepancia en la codificación de los datos. Por ejemplo, lo que se observó es que cuando la información era puntual o cuantitativa, la correlación era alta (p.ej., en las categorías sobre el tiempo de escribir en un contexto académico en español o su percepción del nivel de dominio de escritura académica). La Figura 6.4 muestra un ejemplo de la alta correlación en la categoría 1: tiempo en México. En esta figura se puede apreciar la correlación (Kappa de Cohen) en cada uno de los entrevistados. Como se puede ver, en casi todos los casos ésta es igual a 1 (correlación exacta).

Figura 6.4: Ejemplo de categoría con altas correlaciones entre codificadores

Node	Source	Source Folder	Source Size	Kappa
00-Tiempo en Méxi	01-06	Internals\\Transcription	10457 chars	0.4988
00-Tiempo en Méxi	02-016	Internals\\Transcription	9400 chars	0.9443
00-Tiempo en Méxi	03-077	Internals\\Transcription	7872 chars	0.9332
00-Tiempo en Méxi	04-073	Internals\\Transcription	6309 chars	0.9333
00-Tiempo en Méxi	05-075	Internals\\Transcription	5240 chars	1
00-Tiempo en Méxi	06-085	Internals\\Transcription	6043 chars	0.7402
00-Tiempo en Méxi	07-08	Internals\\Transcription	26659 chars	0.7015
00-Tiempo en Méxi	08-066	Internals\\Transcription	12229 chars	1
00-Tiempo en Méxi	09-078	Internals\\Transcription	8633 chars	0.9677
00-Tiempo en Méxi	10-067	Internals\\Transcription	10884 chars	1
00-Tiempo en Méxi	11-069	Internals\\Transcription	12809 chars	1
00-Tiempo en Méxi	12-070	Internals\\Transcription	8994 chars	0.9473
00-Tiempo en Méxi	13-076	Internals\\Transcription	6246 chars	0.9523
00-Tiempo en Méxi	14-079	Internals\\Transcription	15431 chars	1
00-Tiempo en Méxi	15-020	Internals\\Transcription	9894 chars	1
00-Tiempo en Méxi	16-015	Internals\\Transcription	11394 chars	0.9599
00-Tiempo en Méxi	17-084	Internals\\Transcription	12786 chars	0.9333
00-Tiempo en Méxi	18-086	Internals\\Transcription	12560 chars	0.9545
00-Tiempo en Méxi	19-080	Internals\\Transcription	10735 chars	1
00-Tiempo en Méxi	20-04	Internals\\Transcription	26703 chars	1
00-Tiempo en Méxi	21-068	Internals\\Transcription	6852 chars	1
00-Tiempo en Méxi	22-082	Internals\\Transcription	7992 chars	1
00-Tiempo en Méxi	23-083	Internals\\Transcription	14558 chars	1
00-Tiempo en Méxi	24-072	Internals\\Transcription	10931 chars	0.96
00-Tiempo en Méxi	25-074	Internals\\Transcription	9606 chars	1
01-Tiempo de escrit	01-06	Internals\\Transcription	10457 chars	0.9756
01-Tiempo de escrit	02-016	Internals\\Transcription	9400 chars	0.8567

Sin embargo, cuando la información era repetida en varias ocasiones, la respuesta era muy amplia o no se mencionaba, entonces la correlación no era tan alta. En el primer caso, el problema radicaba en no saber qué información codificar cuando se repetía en varias ocasiones, por lo que se decidió solamente elegir la primera vez que se mencionaba dicha información. En el caso de información que era muy amplia, como en la categoría del proceso de escritura académica, se optó por clasificar toda la información que fuera necesaria para comprender la idea del proceso de escritura. En el caso de información que no se mencionaba, se optó por codificarla. Por ejemplo, en la categoría que hacía alusión a la experiencia con exámenes escritos, la mayoría mencionó que no contaba con experiencia; sin embargo, uno de los codificadores no estaba registrando dicha información. La Figura 6.5 presenta el caso de la categoría 10, experiencia con exámenes escritos, en donde se percibieron bajas correlaciones (la mayoría igual a 0).

Figura 6.5: Ejemplo de categoría con bajas correlaciones entre codificadores

14-Entrevistas estudiantes posgrado.nvp - NVivo P

FILE HOME CREATE DATA ANALYZE QUERY EXPLORE LAYOUT VIEW

Go Refresh Open Properties Edit Paste Copy Merge Cut Copy Merge Format Paragraph Styles Editing

Workspace Item Clipboard Paragraph Styles Editing

Queries Look for Search In Queries Find Now Clear Advanced Find

Queries Results

Queries

Comparación junio 2016 Result

Name	Node	Source	Source Folder	Source Size	Kappa
Comparación junio 2016	10-Experiencia y dif	03-077	Internals\Transcription	7872 chars	0
Pilotaje con 4 entrevistas	10-Experiencia y dif	04-073	Internals\Transcription	6309 chars	1
	10-Experiencia y dif	05-075	Internals\Transcription	5240 chars	0
	10-Experiencia y dif	06-085	Internals\Transcription	6043 chars	1
	10-Experiencia y dif	07-08	Internals\Transcription	26659 chars	0.885
	10-Experiencia y dif	08-066	Internals\Transcription	12229 chars	0
	10-Experiencia y dif	09-078	Internals\Transcription	8633 chars	0.2232
	10-Experiencia y dif	10-067	Internals\Transcription	10884 chars	0
	10-Experiencia y dif	11-069	Internals\Transcription	12809 chars	0.7182
	10-Experiencia y dif	12-070	Internals\Transcription	8994 chars	0
	10-Experiencia y dif	13-076	Internals\Transcription	6246 chars	0
	10-Experiencia y dif	14-079	Internals\Transcription	15431 chars	0
	10-Experiencia y dif	15-020	Internals\Transcription	9894 chars	0.9811
	10-Experiencia y dif	16-015	Internals\Transcription	11394 chars	0.9818
	10-Experiencia y dif	17-084	Internals\Transcription	12786 chars	0
	10-Experiencia y dif	18-086	Internals\Transcription	12560 chars	0.4009
	10-Experiencia y dif	19-080	Internals\Transcription	10735 chars	0
	10-Experiencia y dif	20-04	Internals\Transcription	26703 chars	0
	10-Experiencia y dif	21-068	Internals\Transcription	6852 chars	0
	10-Experiencia y dif	22-082	Internals\Transcription	7992 chars	1
	10-Experiencia y dif	23-083	Internals\Transcription	14558 chars	1
	10-Experiencia y dif	24-072	Internals\Transcription	10931 chars	1
	10-Experiencia y dif	25-074	Internals\Transcription	9606 chars	0
	11-Experiencia y dif	01-06	Internals\Transcription	10457 chars	1
	11-Experiencia y dif	02-016	Internals\Transcription	9400 chars	-0.0129
	11-Experiencia y dif	03-077	Internals\Transcription	7872 chars	0.741
	11-Experiencia y dif	04-073	Internals\Transcription	6309 chars	0

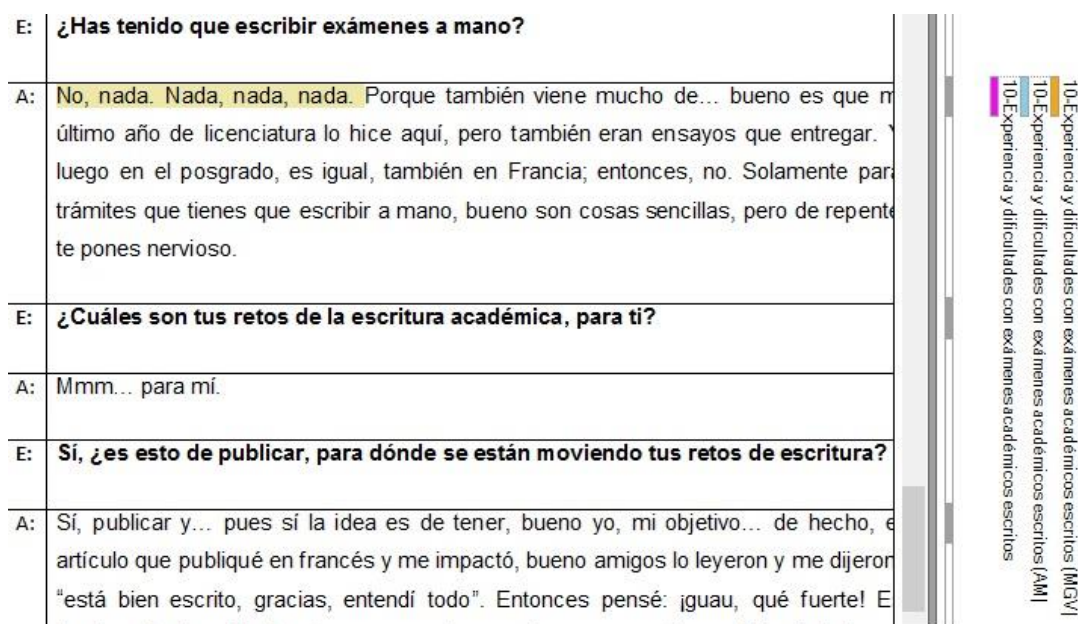
Sources

Nodes

Classifications

Las Figuras 6.6, 6.7 y 6.8 muestran con detalle la información seleccionada por los dos codificadores. Éstas son de gran utilidad para poder revisar con el colega que está codificando por qué existen discrepancias al momento de codificar. En la Figura 6.6 se observa que los dos codificadores seleccionaron exactamente la misma información (correlación 1). En el centro se aprecia sombreado el texto que selecciono el codificador en una categoría. Del lado derecho se pueden apreciar unos rectángulos que muestran en dónde se está seleccionando la información cada uno de los codificadores.

Figura 6.6: Ejemplo de alta correlación (1) entre codificadores



En la Figura 6.7 se parecía como sólo un codificador seleccionó información (correlación 0). En el centro de la figura se aprecia el texto sombreado: "ya no tengo exámenes". Sin embargo, este texto solamente fue seleccionado por el codificador con iniciales AM, como se muestra del lado derecho de la figura. En este texto vertical se aprecia que se trata de la categoría 10 y al final de la leyenda aparecen las iniciales del codificador. A diferencia de la gráfica anterior, solamente aparece un codificador.

Figura 6.7: Ejemplo de correlación 0 entre codificadores

E:	Ok eh... ¿Tienes exámenes escritos en mesoamericanos, ya tuviste algún examen escrito?	
A:	Ah... ya no tengo exámenes ah... con mis materias ninguna, ah... pero necesito entregar trabajos ah... para cada materia a terminar el semestre todavía quedan pendiente todos trabajos que entregar.	
E:	Ok, muy bien ¿Cuál es tu proceso de escritura? tu eh... Tienes tus ideas, esas ideas normalmente que están en chino, en español y después ¿Cómo las plasmas, cómo las escribes ¿En español, un poco de chino? ¿Qué haces?	
A:	Ajá... a mi parecer me acostumbro a elaborar primero como un gráfico, un grafica con... los conceptos principales y luego desarrollar adaptarte y formarte una forma integral para elaborar luego un como ah... resumen, algo así, y luego voy a desarrollar cada parte concretamente.	
E:	¿Pero eso lo haces todo en español?	
A:	En español.	
E:	¿Cuándo tienes palabras que no conoces o que no sabes exactamente que a... la escribes en chino o qué haces?	

En la Figura 6.8 se aprecia una correlación negativa entre los codificadores. Es decir, el codificador 1: AM, seleccionó una porción de texto diferente a la codificadora 2: GV. En la figura se puede apreciar el contraste entre el texto seleccionado entre ambos codificadores. Esto se observa en el texto sombreado en cada uno de ellos.

Figura 6.8: Ejemplo de correlación negativa entre codificadores

Codificador 1		
E:	¿Aquí puedes escribir en inglés?	
A:	Ahm... No sé, ahm... No, yo creo que los proyectos deben ser en español y bueno, es casi, y los reportes también; porque son destinados a, por ejemplo, autoridades de protección civil a autoridades académicas. No sé si, bueno, supongo que la mayoría leen y entienden el inglés, pero es más de como... de... Mi trabajo es puro en español, las juntas del grupo de trabajo se hacen en español, el trabajo de campo, el idioma vehicular en el Instituto es en español, sí. Y bueno, los artículos que edito, publico en revistas internacionales; en generales son en inglés, casi todas las revistas son en inglés. Pero, los proyectos de investigación para recibir financiamiento son en español.	
E:	Cuando tú escribes ahora en francés, ¿también te cuesta trabajo?	
A:	Escribo regularmente porque... Es diferente porque lo que escribo en francés son más textos de divulgación científica o de... ¿cómo se llama?... <i>travel accounts</i> , ¿ <i>reseoconto</i> ? Ahm... Como historias de viajes, cuentos. Cuando, yo tengo un pequeño blog, entonces cuando me voy a viajar a visitar a Chiapas, yo escribo un pequeño texto para describir lo que vi. Como... Memorias de viaje, sí. Es diferente, no es... Pero para escribir textos científicos lo más fácil, para mí, es en inglés. Y todos los artículos que leo están en inglés, y a veces es difícil traducir algunas palabras en inglés ahm... El desarrollo de la ciencia que trabajo es en inglés.	

Codificador 2

E:	¿Aquí puedes escribir en inglés?	
A:	Ahm... No sé, ahm... No, yo creo que los proyectos deben ser en español y bueno, es casi, y los reportes también; porque son destinados a, por ejemplo, autoridades de protección civil a autoridades académicas. No sé si, bueno, supongo que la mayoría leen y entienden el inglés, pero es más de como... de... Mi trabajo es puro en español, las juntas del grupo de trabajo se hacen en español, el trabajo de campo, el idioma vehicular en el Instituto es en español, sí. Y bueno, los artículos que edito, publico en revistas internacionales; en generales son en inglés, casi todas las revistas son en inglés. Pero, los proyectos de investigación para recibir financiamiento son en español.	ing Density 11-Experiencia y dificultades con trabajos académicos (M/GV) 11-Experiencia y dificultades con trabajos académicos
E:	Cuando tú escribes ahora en francés, ¿también te cuesta trabajo?	
A:	Escribo regularmente porque... Es diferente porque lo que escribo en francés son más textos de divulgación científica o de... ¿cómo se llama?... <i>travel accounts</i> , ¿ <i>reseoconto</i> ? Ahm... Como historias de viajes, cuentos. Cuando, yo tengo un pequeño blog, entonces cuando me voy a viajar a visitar a Chiapas, yo escribo un pequeño texto para describir lo que vi. Como... Memorias de viaje, sí. Es diferente, no es... Pero para escribir textos científicos lo más fácil, para mí, es en inglés. Y todos los artículos que leo están en inglés, y a veces es difícil traducir algunas palabras en inglés ahm... El desarrollo de la ciencia que trabajo es en inglés.	

Con esta valiosa información, tras el proceso de codificación del investigador y la especialista, se pudo hacer una revisión minuciosa de aquellas categorías que presentaban bajas correlaciones. Una vez que se revisaron estas discrepancias, se hizo una segunda revisión de las codificaciones por parte del investigador y la especialista.

6.4. RESULTADOS

En este apartado se reportan los resultados del estudio cuantitativo y cualitativo. En primer lugar, se reportan los datos cuantitativos de la correlación del puntaje obtenido en tres instrumentos: la PE del EXELEAA, la encuesta de autoevaluación y la entrevista semiestructurada. Posteriormente, se reportan los resultados del análisis cualitativo de los tres grupos de estudiantes: nivel intermedio bajo, intermedio alto y avanzado.

6.4.1. PRIMERA PARTE: LA CORRELACIÓN DE LA PE DEL EXELEAA CON OTROS INSTRUMENTOS DE EVALUACIÓN DE ESCRITURA ACADÉMICA

Una vez que se evaluó a los 100 examinados, de acuerdo con el resultado que obtuvieron en la prueba escrita del EXELEAA⁷, se los dividió en cuatro grupos: bajo, intermedio bajo, intermedio alto y avanzado. La forma de dividir los grupos fue mediante cuartiles; es decir, del 0 al 3, el primer cuarto se clasificó como bajo, el segundo cuarto como intermedio bajo, el tercer cuarto como intermedio alto y el cuarto como avanzado. El Cuadro 6.1 muestra los

⁷ Recordemos que para poder jerarquizar a los estudiantes, se tomó el puntaje justo arrojado por el programa FACETS.

cuatro bloques en los cuales fueron agrupados los examinados de acuerdo con el resultado en la PE del EXELEAA. Como se puede apreciar, los grupos inferiores (nivel bajo y nivel intermedio bajo) corresponden en su mayoría, en porcentajes, a estudiantes de licenciatura (30.9% de licenciatura vs 4.5% de posgrado), mientras que los superiores (nivel intermedio alto y nivel avanzado), a los de posgrado (95.55% de posgrado vs 69.08% de licenciatura). En el Anexo 3.5 se presentan los niveles de logro de cada uno de los cuatro grupos antes descritos.

Cuadro 6.1: Distribución de estudiantes en grupos de acuerdo con el puntaje obtenido en la PE del EXELEAA

Alumno	Puntaje medio (establecido por FACETS)	Nivel de estudio de los estudiantes	Porcentaje	Nivel
47,61	De 0 a 0.75	2 de grado 0 de posgrado	3.63% grado 0% posgrado	bajo
12,63,49,50,29,91,35,92,88,89,45,90,43,25,23,54,99	De 0.76 a 1.5	15 de grado 2 de posgrado	27.27% grado 4.4% posgrado	intermedio bajo
60,40,52,55,41,58,62,14,95,26,30,27,64,87,59,96,39,73,42,31,21,56,100,98,65,24,94,13,38,16,34,19,6,71,51,77,79,85,75,48,93,18,17,44	De 1.51 a 2.25	28 de grado 16 de posgrado	50.90% grado 35.55% posgrado	intermedio alto
68,28,76,3,66,46,22,36,78,4,32,72,20,10,69,33,84,74,7,83,97,1,5,15,70,57,37,8,11,2,86,82,80,67,53,81,9 ⁸	De 2.26 a 3	10 de grado 27 de posgrado	18.18% grado 60% posgrado	avanzado

La encuesta fue respondida por 22 estudiantes y 25 estudiantes participaron en la entrevista semiestructurada. Sin embargo, no todos los que contestaron la encuesta participaron en la entrevista y viceversa. Una vez que se transcribieron las entrevistas, éstas fueron evaluadas mediante la rúbrica analítica diseñada específicamente para tal fin. Debido a que no siempre se contaba con toda la información que la rúbrica evaluaba, en estos casos, los evaluadores asignaron el valor NA (no aplica). Al final de cada evaluación, se obtuvo la media de los puntajes otorgados en cada una de las 12 categorías de la rúbrica analítica. La correlación de los dos evaluadores de la entrevista fue de 0.86, lo cual demostró que la jerarquización de los estudiantes de acuerdo con su nivel de escritura fue muy similar. En caso

⁸ En color claro se destaca a los estudiantes de posgrado que participaron en la entrevista semiestructurada y/o en la encuesta en línea.

de discrepancia de más de 10%, las entrevistas fueron evaluadas por la especialista en evaluación cualitativa. En estos casos, se sacó el promedio de las tres evaluaciones. En total, 8 entrevistas de 25 tuvieron que ser reevaluadas, y solo una de ellas presentó una discrepancia ligeramente mayor al 20%. El Cuadro 6.2 muestra los resultados de la evaluación de la entrevista semiestructurada, en un rango de 0 a 3, equivalente a la media de las 12 categorías evaluadas por la rúbrica analítica. Los datos se encuentran ordenados de menor a mayor de acuerdo con el resultado que obtuvieron los estudiantes en la entrevista.

Cuadro 6.2: Distribución de estudiantes en grupos de acuerdo con el puntaje medio obtenido en la entrevista

Examinado	Evaluador 1	Evaluador 2	Evaluador 3	Promedio de evaluadores
6	0.82	0.83		0.83
16	1.38	0.80	1.43	1.20
77	1.09	1.58	1.18	1.29
73	1.36	1.22		1.29
75	1.30	1.56		1.43
85	1.33	2.33	1.89	1.85
8	2.00	2.11		2.06
66	2.10	2.08		2.09
78	1.80	2.30	2.30	2.13
67	2.00	2.36	2.20	2.19
69	2.20	2.18		2.19
70	1.91	2.45	2.30	2.22
76	2.20	2.30		2.25
79	2.45	2.18		2.32
20	2.45	2.40		2.43
15	2.40	2.60		2.50
84	2.40	2.67		2.53
86	2.50	2.60		2.55
80	2.20	2.73	2.80	2.58
4	2.50	2.67		2.58
68	2.50	2.88	2.57	2.65
82	2.50	2.80		2.65
83	2.60	2.73		2.66
72	2.70	2.91		2.80
74	2.82	3.00		2.91

Una vez que se obtuvieron los datos cuantitativos de los tres instrumentos (PE del EXELEAA, encuesta tipo Likert y la entrevista), se agruparon todos los datos para poder establecer las correlaciones entre éstos. El Cuadro 6.3 muestra los datos cuantitativos de los estudiantes que participaron en el examen, en la entrevista y/o la encuesta. Los estudiantes se encuentran jerarquizados de menor a mayor de acuerdo con el puntaje medio que obtuvieron en la PE del EXELEAA (segunda columna). Recordemos que la PE del EXELEAA se puntuó con una rúbrica analítica en un rango de 0 al 3; la encuesta tipo Likert contaba con valores del 0 al 5; y la entrevista semiestructurada se evaluó con una rúbrica analítica también con niveles del 0 al 3. En ninguno de los tres instrumentos se ubicó a ningún estudiante dentro del grupo bajo, en color naranja se observan los estudiantes clasificados en el grupo intermedio bajo. En amarillo se observar los clasificados en el grupo intermedio alto, y en verde los que figuran dentro del grupo alto.

Cuadro 6.3: Resultados de los puntajes medios obtenidos en la PE del EXELEAA, la encuesta y la entrevista

Número del examinado	Resultado de la PE del EXELEAA Rango (0-3)	Resultado de la encuesta Rango (0-5)	Resultado de la entrevista Rango (0-3)
73	1.78		1.29
65	1.91	3.45	
16	2		1.2
6	2.03	2.82	0.83
71	2.03	2.82	
77	2.07		1.29
79	2.13		2.32
85	2.13	3.18	1.85
75	2.15	3.39	1.43
68	2.26	3.88	2.65
76	2.32		2.25
66	2.34	3.06	2.09
78	2.38		2.13
4	2.39	3.67	2.58
72	2.42	3.88	2.8
20	2.43	3.55	2.43
69	2.44	3.70	2.19
84	2.47		2.53
74	2.51	3.64	2.91
83	2.55	3.79	2.66

1	2.58	4.73	
15	2.58	3.55	2.5
70	2.61		2.22
8	2.65	3.88	2.06
11	2.67	3.58	
86	2.71	4.70	2.55
82	2.75	3.27	2.65
80	2.77		2.58
67	2.8	4.55	2.19
81	2.86	2.91	
9	2.87	4.27	
MEDIA GRUPAL	2.406	3.65	2.167

Con los resultados obtenidos por cada uno de los estudiantes, se establecieron las correlaciones entre cada uno de los grupos en SPSS. La Tabla 6.1 muestra las correlaciones entre los tres instrumentos. Como podemos observar, las correlaciones entre el examen y la encuesta (0.507), y entre la encuesta y la entrevista (0.494) son moderadas, mientras que la correlación entre el examen y la entrevista (0.718) es fuerte. Naturalmente, debido a la amplitud de la muestra, los datos no pueden ser generalizables; no obstante, nos aportan valiosa información sobre la estrecha relación que existe entre el desempeño de los candidatos en una PE en un ECLAA, en comparación con su dominio de escritura académica.

Tabla 6.1: Correlaciones de Pearson (examen, encuesta y entrevista)

		Examen	Encuesta	Entrevista
Examen	Correlación de Pearson	1	.507*	.718**
	Sig. (bilateral)		.016	.000
	N	31	22	25
Cuestionario	Correlación de Pearson	.507*	1	.494
	Sig. (bilateral)	.016		.052
	N	22	22	16
Entrevista	Correlación de Pearson	.718**	.494	1
	Sig. (bilateral)	.000	.052	
	N	25	16	25
*. La correlación es significativa al nivel 0,05 (bilateral).				
**. La correlación es significativa al nivel 0,01 (bilateral).				

6.4.2. SEGUNDA PARTE: LA ESCRITURA ACADÉMICA DE LOS ESTUDIANTES NO HISPANOHABLANTES

La segunda parte de este estudio consistió en identificar qué patrones o características en común presentaban los estudiantes de posgrado, en cuanto a su escritura académica, que participaron en la entrevista. Para poder hacer esto, se dividió a los estudiantes en cuartiles dependiendo del resultado que obtuvieron en la evaluación de la entrevista. Ningún estudiante fue clasificado dentro del grupo bajo. El cuadro 6.4 muestra la organización de los estudiantes en los siguientes tres grupos: intermedio bajo, intermedio alto, avanzado.

Cuadro 6.4: Estudiantes clasificados en grupos de acuerdo al puntaje que obtuvieron en la entrevista

Grupo 1	Estudiante	Grupo 2	Estudiante	Grupo 3	Estudiante
Intermedio bajo (0.76 a 1.50)	6	Intermedio alto (1.51 a 2.25)	85	Avanzado (2.26 a 3.00)	79
	16		8		20
	77		66		15
	73		78		84
	75		67		86
	75		69		80
			70		4
			76		68
					82
					83
			72		
			74		

A continuación, se hará una descripción detallada de la información recabada en cada uno de los grupos y de acuerdo con las categorías de análisis previamente identificadas. También se presentan algunos ejemplos representativos de la información proporcionada por los estudiantes. El orden en que se presentarán los comentarios de los estudiantes corresponderá a cómo fueron puntuados en la entrevista; es decir, se comenzará por los más bajos de cada grupo. De las 19 categorías que se presentarán, solamente la 6 fue de carácter emergente. Esta categoría hacía alusión a la necesidad de tomar o no cursos de lengua. Al final de la descripción de cada uno de los grupos, también se presenta información de la categoría 20, la cual incluía otros conceptos no considerados en el resto de las categorías. Las categorías que mostraban información comparable entre sí se presentan en dos columnas. Las categorías 16 (uso de recursos materiales para la escritura) y 17 (edición con ayuda de terceros) se fusionaron con la 15 (proceso de escritura), pues se encontraban íntimamente relacionadas entre sí, además de que su integración facilitaba la comprensión del proceso de escritura.

6.4.2.1. GRUPO INTERMEDIO BAJO

En el grupo intermedio bajo fueron identificados 5 estudiantes: dos de doctorado, dos de maestría recién llegados a México y uno realizando una estancia postdoctoral. En relación con el tiempo que tenían de vivir en México y de escribir en un contexto académico, se observa que éste era muy semejante, aunque destaca que tres de ellos ya tenían aproximadamente un año escribiendo en un contexto académico; no obstante, sus habilidades de escritura no eran aún suficientes.

Categoría 1. Tiempo de vivir en México**E6:** Un año y medio**E16:** Un año y tres meses**E77:** Casi tres años**E73:** Cuatro meses**E75:** Cuatro meses**Categoría 2. Tiempo de escribir en un contexto académico en español****E6:** Poco menos de un año**E16:** Cinco meses, más o menos**E77:** Dos años y medio**E73:** Cuatro meses**E75:** Cuatro meses

En cuanto a su percepción del nivel de español, reportaron que éste era bastante malo, muy malo, malo, suficiente o no tan bueno. De igual manera, sus sensaciones al momento de escribir fueron negativas: frustración, carencia de valor en lo escrito, tristeza, sentirse limitado, o percibirse como tonto. Sólo uno mencionó que se sentía con más confianza; sin embargo, este estudiante tenía apenas cuatro meses en México y poca experiencia de escribir en un contexto académico.

Categoría 3. Nivel de escritura en español**E6:** Entre malo y suficiente**E16:** Bastante malo**E77:** Muy mal**E73:** No es tan bueno**E75:** Creo que suficiente**Categoría 4. Sensaciones al momento de escribir en español****E6:** [...] estoy muy triste, muy frustrada. Siento que no puedo expresarme [quiso decir expresarme]. [...] No puedo decir lo que quiero decir. Es como si no pudiera mostrar mi valor real.**E16:** Un poco, a veces un poco así, limitado. Sí, limitado, pero no es una frustración ni una ansiedad.**E77:** Ah... triste (risas). Me siento muy triste, me

siento muy tonta.

Investigador: ¿Pero te sientes con confianza cuando tienes que escribir en español?

E73: Mmm... No, todavía no (risas).

E75: Ya con más confianza, porque ya estoy aquí unos meses

En relación con el tiempo que les tomaba escribir en español, los comentarios fueron que les tomaba mucho o muchísimo más que en la lengua materna, más o menos dos veces el tiempo. Destaca que uno de ellos mencionó que aunque invertía más tiempo, eso estaba bien, pues era importante adquirir esa habilidad.

Categoría 5. Tiempo que toma escribir en español

E6: Pues mucho más que en mi lengua, pues me cuesta mucho más trabajo. Dos veces, tal vez, el tiempo.

E16: [...] me cuesta casi dos veces más tiempo [...] Nunca es demasiado; porque lo más que escribo, lo más que progreso y... también escribo para cosas importantes para cual nunca se invierte demasiado tiempo.

E77: Muchísimo más.

E73: [...] necesito como mucho eh... ehm... más tiempo que normalmente en alemán o en inglés. [...] El doble.

E75: Es un poco más sí, claro que escribo más lento que puedo escribir que en alemán, pero... está bien. [...] 50% más, depende.

Sobre la opinión o necesidad de tomar cursos de lengua, reconocieron en general que eran importantes, siempre y cuando éstos se encontraran alineados a sus requerimientos académicos; no obstante, hubo quienes afirmaron que también las clases de lengua eran importantes y quizás aún más que las del posgrado.

Categoría 6. Toma clases de español o manifiesta una opinión al respecto

E6: [...] Yo creo que sería bien si los extranjeros tuvieran curso de español en su entidad en el posgrado. Creo que estaría una buena cosa.

Investigador: ¿Alguna vez viniste a tomar cursos aquí [refiriéndose al CEPE]?

E6: Sí, en el grupo básico cuatro, pero no me gustó. Por nada. Para mí era súper lento, súper, súper lento. No sé si es por la maestra, pero no tenía ninguna gana de dar clase.

E77: [...] quiero tomar clase, pero me da flojera. [...] yo vivó poco lejos. Y... El caso de escribir, ajá. De hecho no es importante la clase, revisar es más importante. Yo pienso que si eh... No necesita clase, como sólo escribir, y después si podemos... Corregir, quizás como la clase privada, sólo para una persona. Quizás es mejor.

E73: Estudi...é como un año en la universidad en Alemania [refiriéndose al español] y aquí como... seis meses en la CEPE [es el tiempo estudiando español en el Centro de Enseñanza para Extranjeros].

Investigador: ¿Y los cursos del CEPE te ayudan en eso [escritura académica], o no específicamente?

E73: No específicamente. Es más solo sobre el gramático y no sobre mis, como temáticos geografía.

E75: [...] Tenía 3 clases, pero... cuando empezó, empecé con mi curso de español aquí en el CEPE, no tenía tiempo para dos de las clases de geografía. [...] A mí me parece más importante, antes de estudiar, el aprender español y mejorar mi español.

En cuanto a las clases en el posgrado, los estudiantes de doctorado mencionaron que no eran obligatorias, pero que asistían como oyentes. Los estudiantes de maestría dijeron que sí tomaban clases en el posgrado.

Categoría 7. Clases en el posgrado

E6: No tengo clase.

E77: ... De hecho el curso de doctorado no hay clase; pero estoy tomando clase por oyente, de la clase de universitaria.

E73: Ajá, tengo dos clases.

E75: [...] Tenía 3 clases, pero... cuando empezó, empecé con mi curso de español aquí en el CEPE, no tenía tiempo para dos de las clases de geografía. Entonces, solo tengo un curso de geografía ahora, pero no es tan grave, porque este curso es el único que van aceptar en Alemania, y los otros dos no eran tan importante.

Sobre el idioma de las lecturas asignadas por los docentes, se observó que éstas eran mayoritariamente en español. En relación con el idioma de lecturas adicionales, se percibió que el idioma de éstas se encontraba condicionado por el área de estudio.

Categoría 8. Idioma de las lecturas asignadas

E77: Español

Categoría 9. Idioma de lecturas adicionales

E6: Siempre en inglés, en español, no. [...] Como el 95%, pero casi todo en inglés.

E73: Español

E75: Es todo en español, a veces en inglés.

E75: Sería en español, porque... los temas son de aquí, entonces la escritura también es en mexicana... en español.

En cuanto al idioma de las notas de clase, se percibió una fuerte influencia de la lengua materna para tomar notas y escribir información importante para recordar.

Categoría 10. Idioma de las notas de clase

E6: No tengo clase.

E77: Mezcla, pero generalmente escribo coreano porque puede escribir más rápido.

Investigador: ¡Ah, escribes en coreano! Entonces, tú escuchas en español al profesor, pero ¿tú tomas tus notas en coreano?

E77: Sí, en coreano, porque... ah... a veces, yo escribo sólo coreano; pero escribo por coreano, porque letras coreanas, este... ¿Cómo se dice?

Investigador: El alfabeto.

E77: Sí, alfabeto, el coreano este... sonido, letras de sonido; pero si yo escribo en español, a veces yo no puedo saber bien como las letras exactamente; por ejemplo, cómo se escribe como... por ejemplo: *b* o *v*. Como esas cosas, por eso yo escribo por coreano; sólo su sonido, no su sentido.

Investigador: Tú escribes en español usando el alfabeto coreano.

E77: Ah... pero no todo. Yo escribo por coreano; pero solo las palabras, palabras importantes. Si yo no sé su sentido, yo escribo su sonido.

E73: La mejor, mejoridad (quiso decir mayoría) son en... Español, porque el profesor hable en español. Entonces, es más fácil sólo escribir en español, cuando no sé traducir, pero son dos. Cada vez es cuando yo sólo escribía la idea, escribe también cada vez es en alemán, porque la idea es más... Fácil en alemán y más corto que... Para mi escribir así.

E75: Depende, la mayoría pienso que lo hago en español, pero a veces también en alemán.

Sobre la experiencia con exámenes escritos, ésta fue nula en todos los casos. En el caso de los dos estudiantes de maestría, apenas estaban terminando su primer semestre académico. Los demás estudiantes eran de doctorado, donde se evidenció que los exámenes a este nivel no son frecuentes. En cuanto a los trabajos escritos, los estudiantes de maestría aún no presentaban trabajos finales. Los trabajos escritos a lo largo del semestre habían sido cortos. Los estudiantes de doctorado habían redactado unidades temáticas o capítulos de tesis. El postdoctorando reportó que la mayoría de sus escritos eran para proyectos académicos de la UNAM, pero que sus publicaciones eran en inglés.

Categoría 11. Experiencia con exámenes escritos

E6: Tengo exámenes, pero nada más oral.

E77: Examen no, no , no. Nunca.

E75: No, hasta ahora no.

Categoría 12. Experiencia con trabajos escritos

E16: Los proyectos deben ser en español y bueno, es casi, y los reportes también [...] bueno, los artículos que edito, publico en revistas internacionales; en generales son en inglés, casi todas las revistas son en inglés. Pero, los proyectos de investigación para recibir financiamiento son en español. [...] Para escribir textos científicos lo más fácil, para mí, es en inglés.

E77: [...] Un capítulo de mi tesis.

E73: [...] No necesité trabajar mucho, sólo a casa. Hasta ahora sólo necesito leer

E75: [...] Este fin de semestre tengo que escribir un texto sobre un caso de desastres y ese va ser mi primero texto que es un poco más largo, primero solo eran como tareas o textos muy cortos, pero nada más...

En cuanto a las dificultades al escribir en español, destacaron varios problemas. En primer lugar, se hizo notoria la dificultad de escribir con un estilo académico formal; en segundo lugar, se percibió el proceso de escritura a modo de traducción entre la lengua materna y el español; finalmente, se observó una dependencia al diccionario. Los problemas de escritura se evidenciaron, sobre todo, a nivel oracional, gramatical y léxico.

Categoría 13. Dificultades al escribir en español

E06: El estilo. Casi siempre yo pienso en francés y luego hago una traducción casi palabra por palabra. [...] Para mí es más a nivel de una oración.

E16: [...] yo diría la gramática, y la ortografía, y también la... no sé cómo explicarlo, pero el modo de... ¡el estilo! que... cuando hablas francés, tienes... la estructura de oraciones se conserva cuando escribes en español. [...] Ahm... Conjugación los verbos (risas).

E77: Ah... Cuando yo escribo, no puedo escribir en una vez, porque tengo que checar varias veces diccionario e internet. Y sin internet, sin diccionario, yo no puedo acabar escribir.

E73: Creo que más las palabras mmm... Y también yo, entiendo más, pero escribir en... buen español es... no tan fácil para mí todavía.

E75: [...] Depende, es que me faltan... a veces muchas palabras, pero con un diccionario eso no es un problema y la otra cosa es la gramática que... a veces me hace unos problemas.

En relación con la lengua materna, se mencionó que las lenguas cercanas al español ayudaban al proceso de escritura, y también reconocieron que lenguas distantes al español no ayudaban al proceso de escritura en esta lengua.

Categoría 14. Ayuda de la lengua materna para escribir en español

E16 (belga): [...] yo creo que el estilo... No hay una diferencia tan grande de estilo entre el francés y el español. Son diferencias ligeras ¿no?, por ejemplo, en el japonés o en el ruso, sería mucho más... [distante]

E77 (surcoreana): Nooooo, nooo. Nada, (risas).

Investigador: ¿La estructura?, ¿la forma?...

E77: Totalmente diferente.

Investigador: No te ayuda.

E77: ¡No me ayuda nada!

En relación con el proceso de escritura, se decidió fusionar esta categoría con el uso de recursos y la ayuda de terceros para editar y revisar sus textos, pues las tres se encontraban íntimamente relacionadas. En estas categorías, se evidenciaron las dificultades que los estudiantes tienen para escribir en español, así como la fuerte dependencia a diccionarios, traductores electrónicos o foros de discusión, aunado a la dificultad de poder discernir entre las distintas acepciones de una palabra. En general, el proceso de escritura se describió como una traducción de la lengua materna al español, o bien escriben en español, pero la estructura del pensamiento se encuentra en otra lengua. Esta forma de escritura, según comentaron, interrumpía el flujo del pensamiento por escrito. La incorporación del material proveniente de distintas fuentes era también incipiente. Los dos estudiantes de maestría reportaron que aún no habían escrito trabajos lo suficientemente largos para poder dar cuenta de un proceso de escritura detallado en español. En cuanto al apoyo de terceras personas, cuando el tutor no era mexicano, esto representaba un problema adicional para el estudiante que no contaba con buen nivel de escritura en español. En general, el apoyo de terceras personas era común y muy socorrido por este grupo de estudiantes.

Categorías 15-17. Descripción del proceso de escritura (incluye la fase de lectura, planeación, uso de recursos materiales y de la edición del texto con ayuda de terceros)

E6: Casi siempre yo pienso en francés y luego hago una traducción casi palabra por palabra. [...] para mi segundo semestre, estas unidades... Se dice, pues venían un poco más “natural”.

Investigador: ¿Utilizas diccionarios o traductores?

E6: Sí, sí. Pues no voy directamente a un diccionario, sino a algunos foros donde hay un intercambio entre francés parlantes y mexicanos o español, para tener una idea más general. [...] muchas veces estoy en internet en páginas para aprender el español. Sobre todo, con el subjuntivo, hay después ciertos nexos. Es así que los aprendo, si no conozco un nexo voy a ver en un diccionario, o en un foro...

Investigador: ¿Le pides ayuda a alguien para que revise tus textos?

E6: [...] Pregunté varias veces para que alguien me corrija mis unidades, pero nadie lo hizo y yo creo que es la única cosa que puede ayudarme: conocer mis errores, dónde están mis problemas. Pero, o si no, pues si no, no veo... [...] mi tutor es un ruso [...] Él ya lleva como 18 años aquí, pero todavía hace errores. No es lo mismo que un tutor mexicano. [...] me dijo o me dice: "hay que ver con un hispanohablante". [...] No tengo muchos amigos mexicanos, tengo colegas... del instituto, pero hay como cinco francoparlantes.

E16: [...] Escribo con la mano, a mano [...] y después escribo y... acomodado en la computadora. [...] Es todo en español.

Investigador: ¿Y cuando hay un bloqueo?

E16: [...] Las frases me... Las formulo como lo... como lo siento pero, pero sin ningún otro conector; por eso, el estilo puede ser... mi estilo de escritura español puede ser mal. Cuando es sólo una palabra... Uso mucho *WordReference*, que es bueno porque tienes en general varias opciones y también, tienen foros donde otros usuarios... Enviaron sus dudas y otros más experimentados contestaron esas dudas. En general, esto me ayuda mucho porque hay muchas veces preguntas similares a la mía. Y cuando hay una palabra técnica, que en general no se encuentra nunca, uso *Wikipedia* en francés o en inglés; y después voy a ver el artículo equivalente en español.

Investigador: ¿Todo eso interrumpe tu pensamiento?

E16: Sí, sí, sí. Esto sí, se cortó el hilo del pensamiento.

E77: Ah, primero escribo... como en coreano, esquema y después escribo, y como... explicar, explico en coreano. Escribo otra vez y después escribo en español.

Investigador: Entonces, haces un ejercicio de traducción del coreano al español.

E77: Sí, podemos decirlo así. [...] Es problema de mi manera de hablar. [...] En coreano yo puedo hablar así como nivel alto, sí, pero en español yo no puedo hablar así. Mi oración es, parece como de niño de secundaria, por eso no es tan... como educativo, ¿ajá? Y también es problema de... palabras, porque yo uso diccionario de coreano y español, español y coreano. En mi diccionario dice: Esa palabra significa esto, eso, eso, pero cuando yo escribo mi tesis, de hecho esa palabra no se dice ese sentido, ¿ajá? Por eso siempre yo necesito alguien puede leer otra vez. [...] También yo uso la oración de los libros; este libro una oración, este libro una oración. Por eso un párrafo puede ser, no tener mucha ilación.

Investigador: ¿Utilizas otros recursos?

E77: Sí, todo: diccionario, también *translator* de google; sólo google, también yo escribo una oración y escribo en google, para que... saber: ¿las personas están usando esa oración, o no? ¿Esa palabra, o no?

Investigador: ¿Le pides ayuda a alguien para que revise tus textos?

E77: Yo... reviso con mis compañeros primero. [...] Mexicanos. Sólo mexicanos, ajá. Porque ellos tienen conocimiento de historia, necesito ese tipo de personas. Pues primero, yo lo veo con mis compañeros, y después yo ah.... Lo revisamos juntos con otros amigos mexicanos, que puedan hablar o leer bien español, pero que no tienen conocimiento de historia. Lo hago dos veces.

En cuanto a la retroalimentación que recibían de sus escritos, dado que los estudiantes de maestría aún estaban por escribir sus trabajos finales, no reportaron nada al respecto. Los dos de doctorado mencionaron que era limitada.

Categoría 18: Retroalimentación de la lengua

E6: [...] Pregunté varias veces para que alguien me corrija mis unidades, pero nadie lo hizo y yo creo que es la única cosa que puede ayudarme: conocer mis errores, dónde están mis problemas. Pero, o si no, pues si no, no veo... [...] mi tutor es un ruso [...] Él ya lleva como 18 años aquí, pero todavía hace errores. No es lo mismo que un tutor mexicano. [...] me dijo o me dice: "hay que ver con un hispanohablante". [...] No tengo muchos amigos mexicanos, tengo colegas... del instituto, pero hay como cinco francoparlantes.

E77: Primero contenido, pero después español, pero no tan...

Investigador: ¿Profundo?

E77: No tan profundo, porque todavía mi capítulo están cambiando, sigue cambiando. Por eso.

Investigador: Entonces, está la preocupación por el contenido.

E 77: Y además, cuando mi profesor lee mi capítulo, casi no hay muchos problemas; porque ya revisamos tres veces con mis amigos, de hecho, ajá.

Finalmente, sobre qué los ayudaría a mejorar su escritura académica, los estudiantes afirmaron que la corrección individual ayudaría significativamente. También enfatizaron que la lectura, la práctica y cursos de lengua en el posgrado podrían fomentar el desarrollo de esta habilidad.

Categoría 19. Reflexión sobre qué podría mejorar su escritura académica

E6: [...] Pregunté varias veces para que alguien me corrija mis unidades, pero nadie lo hizo y yo creo que es la única cosa que puede ayudarme: conocer mis errores, dónde están mis problemas. [...] Yo creo que sería bien si los extranjeros tuvieran curso de español en su entidad en el posgrado. Creo que estaría una buena cosa. [...] Para mí, nada más tomar clase... O leer, o tomar clase. [...] Tomar clase, pero no clase como en el CEPE.

E16: [...] lo más que escribo, lo más que progreso

Investigador: Todo lo que tú lees, me imagino, casi todo es y va a ser en inglés. [...] ¿Crees que si tú leyeras en español sería más fácil escribir en español?

E16: ¡Ah, sí! Claro, claro. Sí, porque hasta en modo inconsciente tú asimilas y memorizas las palabras, los modos de armar las oraciones y todo. Yo tendría que leer más textos científicos en español, sí.

E77: Yo quiero escribir, diario, diario, todos los días. Y también quiero tomar clase, pero me da flojera. [...] Yo pienso que si eh... No necesita clase, como sólo escribir, y después si podemos... Corregir, quizás como la clase privada, sólo para una persona. Quizás es mejor.

E73: [...] yo necesito leer más, desde el primero para la vocabulario y también... escribir pero con escribir, el problema es que necesito algún persona que pueda mmm... corregir.

E75: [...] bueno, tengo a practicar es cierto y también sería bien si alguien quien puede hablar español me ayuda y yo voy a escribir antes y él puede corregir... corregirlo, después para yo pueda ver que faltas yo hago.

6.4.2.2. GRUPO INTERMEDIO ALTO

En el grupo intermedio alto se identificó a 8 estudiantes: cuatro de doctorado y cuatro de maestría. Al igual que en el grupo intermedio bajo, el rango de tiempo viviendo en México se extendió desde tres meses hasta cinco años. El rango de tiempo escribiendo en un contexto académico en español fue muy similar, desde cuatro meses hasta cinco años.

Categoría 1. Tiempo de vivir en México

E85: Cinco años

E66: Cuatro años y medio

E78: Dos años y algo

E67: Voy por tres años

E69: Cuatro meses

E76: Tres meses

Categoría 2. Tiempo de escribir en un contexto académico

E8: Como cuatro meses

E66: Hace un año y medio

E78: Como dos años y medio en total

E69: Cinco años

E70: Cuatro meses

E76: Cinco años

Sobre su percepción del nivel de escritura en español, comentaron aspectos tanto negativos como positivos. Dentro de los negativos, mencionaron que es difícil escribir en español, que es la habilidad menos desarrollada o que no es muy académico. Entre las positivas, dijeron que sienten que su nivel está avanzando más, aunque hay que mejorar, que el nivel es satisfactorio y una estudiante mencionó que su nivel se encontraba entre B2 y C1. De las sensaciones al momento de escribir, éstas fueron diversas. Los sentimientos negativos manifestados fueron de terror, de poca confianza o de tener que hacer el esfuerzo por escribir en español. Los sentimientos positivos fueron de gusto, pese a las dificultades de escribir en español o de dudas sobre lo que se escribe.

Categoría 3. Nivel de escritura

E85: Ahm... ahorita está avanzando mi nivel, pero todavía faltan muchos mmm... palabras, y todavía tengo muchos errores.

E66: Sí, es bastante difícil y... eso es lo que me gustaría mejorar.

E67: Yo creo, que de mi español, es lo menos bueno, porque no lo trabajo mucho.

E69: Uy... bueno, en todas las lenguas, mi nivel de escritura creo que siempre es el peor de todos. No es muy académico...

E70: No, bueno mi nivel de escritura sí tiene sus reservas (risas).

E76: Creo que está entre B2 y C1.

Categoría 4. Sensaciones al momento de escribir en español

E85: Traduzco en español, por eso siempre yo utilizo diccionario, siempre, siempre, siempre. Pero está bien me gusta.

E08: Creo que ya, ya es mejor, como el sentimiento que tengo de la parte de escritura es mejor.

E66: ¡Ay!, me aterroriza, porque es como... en general escribir no es algo que me encanta, es...

E67: Sí me gusta, me gusta. No lo veo tan difícil.

E70: Bueno, pues obviamente es algo más difícil, que escribir en mi lengua materna.

Investigador: ¿Pero, te sientes con confianza?

E70: Bueno, digamos que hay que hacerlo, ¿verdad? Hay que escribirlo, así que hago mi mejor esfuerzo.

E76: Mmm... Pues me cuesta un poco mmm... Las conjugaciones, verbos eh.... Pero no es muy difícil.

Investigador: ¿Te sientes con confianza, con seguridad, cuando escribes en español?

E76: Mmm.... No mucho (risas), siempre dudo, ahm... Unas palabras como... las frases (risas).

Cuando se les preguntó sobre el tiempo que les tomaba escribir en español, en general reconocieron que su proceso era un poco más lento tanto en la escritura como en la edición, pero también había a quienes aún les toma un tiempo considerable.

Categoría 5. Tiempo que toma escribir en español

E85: Ahm... ahorita, está disminuyendo el tiempo. Pero antes sí, mucho.

E8: Pues... no tanto, como... pues normal. [...] Al principio, pues fue difícil (risas). [...] Al principio sí que me tomaba demasiado tiempo.

E66: Pienso que el doble.

E67: Diría me cuesta cincuenta por ciento de tiempo más.

E69: [...] Lo que sí tarda un poco más es... formar las ideas en español; porque los textos normalmente están en inglés, por ejemplo. Entonces, eso sí; porque a veces no me sale la palabra alemana y tengo problemas de [no se entiende muy bien, parece que dice traducción] y no sé qué; entonces, en alemán como es lengua materna tengo la lectura y sí, sé cómo escribirlo, ¿no? Pero en español, eso creo que sí toma más, pero ya a la hora de escribir, no es mucho más tiempo.

E70: Pues, creo que en realidad la diferencia no sería tan grande porque lo que a mí más tiempo me lleva es planear como mis ideas, aclarar mis ideas y hacer un concepto de lo que voy a escribir. Digamos que la lengua no es ah... no es tan importante en eso, pues ah... luego más se tarda la parte de editar el texto, eso sí.

En relación con la importancia de tomar clases de lengua, ningún alumno de este grupo mencionó nada al respecto. En cuanto a las materias que tomaban en sus posgrados, todos reportaron estar tomando o haber tomado materias en sus cursos. En general, todos habían tomado varias materias en sus posgrados, algunas de ellas obligatorias y otras no.

Categoría 7. Clases en el posgrado

E85: [...] En mi Facultad no es obligatorio tomar una clase oficial, es opcional. Yo tomé clase el semestre pasado pero, ahorita tengo que avanzar mi tesis.

E8: Pues tomo 3 materias aquí.

E66: Tomé 4 seminarios [en el semestre en curso].

Investigador: ¿Son obligatorias?

E66: No, solamente un seminario fue obligatorio y otros... extracurriculares.

E78: Este semestre eran dos, pero yo tomé tres. Y seminarios.

E69: Yo tomo curso en Estudios Latinoamericanos; un taller así, para la tesis; y otro seminario en Migración.

E76: Sí, sí, sí que tomé dos materiales [quiso decir materias]; pero no, son para Metodología de la Investigación.

En cuanto al idioma de las lecturas que les asignaban en sus cursos, la mayoría comentó que eran español, aunque en algunos casos también tenían que leer textos en inglés. Sin embargo, en cuanto al idioma de las lecturas adicionales en ocasiones éste se encontraba determinado por la investigación, aunque en algunos casos se mencionó que había más material de lectura en inglés.

Categoría 8. Idioma de las lecturas asignadas

E66: Español

E78: Pues español, inglés. Hay algunas cosas en inglés. Hay mucho material en español, por supuesto.

Investigador: Como ¿qué porcentaje estaba en español?

E78: Como la mitad más o menos. Y la otra mitad en inglés. Quizás un artículo en francés. Pero... sí.

E69: En inglés y español.

E70: Bueno, la mayoría era en español, algunos textos eran en inglés. [...] como 15% en inglés; 85% en español, por ahí.

E76: Sí, en español. Sí, algunos documentos en inglés, que eh... la profesora dio a leernos pero... muy, muy pocos.

E69: Uff... Este... En inglés y español, sí. Así me los asignan [...]

Categoría 9. Idioma de lecturas adicionales

E85: [...] es que mi investigación es como influencia de... Las culturas entre ambas países, de Japón y México, por eso tengo que leer en japonés y también en español, y a veces en inglés.

Investigador: ¿Qué porcentaje dirías tú que lees en español? ¿Qué porcentaje en japonés?

E85: Mmm... Depende del capítulo, como si yo escribo sobre la cultura japonesa, casi no hay... literatura ni bibliografía en español, por eso leo mucho en japonés, y después traduzco en español. Pero, sobre la cultura mexicana en Japón casi no hay libros, por eso tengo que leer los libros de aquí.

E66: En inglés, polaco, sobre todo. Sobre todo, en inglés, porque también es lo más accesible en el internet.

E78: Pues italiano casi no hay producción. Y antropología no hay producción. Entonces... y por lo que hago yo eh... libros en español que hablan de antropología tienen una caracterización teórica, no todos por supuesto, pero muchos, un poco vieja. Y entonces, eh... pues, no elijo por la lengua, más bien por el contenido. Y sí, la mayoría en inglés.

E67: Eso siempre está en inglés. Casi 100%

E69: [...] me gustan más lo del inglés. Leer sí, toma mucho más tiempo, leyendo este... español académico. [...] la lectura española aún me cuesta, porque lo leo y las frases a veces son así de largas, como en alemán, que nosotros también escribimos muy así, no es muy comprensible (risas).

E70: Yo las elijo en español. Yo considero que entiendo mejor el español que el inglés.

En relación con las notas de clase, hubo quienes mencionaron mezclar entre español y otra lengua; sin embargo, hay quienes reconocieron escribir mayoritariamente en español y sólo eventualmente en otro idioma para facilitar la escritura.

Categoría 10. Idioma de las notas de clase

E66: ¡Ay! Es una mezcla terrible. Es una mezcla de todo, porque también, ¿sabes qué? Depende de los temas, porque algunas cosas... si hay cosas que he aprendido en español, por ejemplo, tomo notas en español, hay cosas que he aprendido... en inglés o polaco depende del tema, depende del tema. Sí, y dependiendo del tema, como te comenté, algunos temas más en inglés, algunos en español. Dependiendo de si se relaciona con algún texto que... he leído anterior, había leído anteriormente en inglés o español, eso influye en que idioma tomo las notas.

E78: Uy... (risas) no... ¡Es todo un collage! Es una mezcla entre español, italiano. Pues depende de la lengua en que se está dando. Pues español quizás con algunos conectores en italiano, pero también en las clases por ejemplo, hay también un poco en inglés, pocos conceptos que también están definidos así... o porque simplemente si lees un texto en inglés las notas están en inglés. Y luego, a veces en maya, pero esa es otra cosa.

E76: Normalmente, en español y si algunas veces mmm... No puedo escribir en español (risas), pues lo escribo en chino, lo que entiendo más; por ejemplo, puedo poner solo una, dos caracteres y luego entender.

E67: En aquel entonces [cuando comenzó su doctorado hace tres años] no hablaba tan bien español. Entonces, como me acuerdo que... Mi manera de tomar notas era entender el español y luego escribir en francés; porque se me hacía más fácil, ahora creo que estaría diferente si tuviera que volver a hacerlo.

Investigador: ¿Crees que escribirías en español?

E67: Sí, ya creo que ya lo podría hacer sin problema en español, ya creo que mi cerebro ya está entrenado al español mucho más que antes.

E69: Pues, si estoy en la clase, y estamos hablando en español, pues todo en español, claro. Todo en español (risas). [...] pero normalmente lo que escucho sí lo escribo así. Si se me olvida lo que dijo como... en palabras, que se me olvida la palabra, pues a veces sí lo pongo en inglés o algo.

E76: Normalmente, en español y si algunas veces mmm... no puedo escribir en español (risas), pues lo escribo en chino, lo que entiendo más; por ejemplo, puedo poner solo una, dos caracteres y luego entender.

La gran mayoría asintió que no había exámenes escritos o muy pocos en las clases. Solamente el estudiante de estudios mesoamericanos comentó que había hecho varios exámenes. En cuanto a los trabajos escritos, éstos son mucho más abundantes y de diversa naturaleza. Algunos tienen que escribir capítulos de sus tesis, otros entregar trabajos pequeños durante el semestre, y la gran mayoría presenta trabajos finales largos al final del curso.

Categoría 11. Experiencia con exámenes escritos

E8: Sí, sí, sí. He tenido... no sé, dos o tres exámenes en donde nos pidieron escribir un ensayo de cinco cuartillas, pero de repente, como espontáneamente.

E66: No

E78: ¡Ah! Pues sí, en lengua maya 3. Y quizás uno o dos.

E69: Mmm... Una vez, un examen de media hora así, un examen.

E70: No, no. Aquí, en realidad no se da mucho lo de los exámenes, normalmente son exposiciones o trabajos.

E76: Mmm... aquí no.

Categoría 12. Experiencia con trabajos escritos

E85: [...] tengo que avanzar mi tesis...

E8: [...] hay que entregar tres o cuatro trabajos como de tres cuatro cuartillas y luego viene esta parte del trabajo final que es de quince o veinte páginas.

E66: Para un seminario, teorías de comunicación, tuve que escribir mi trabajo final y sí me costó.

E78: [...] la mayoría son como trabajos que vas a entregar en tal fecha.

E67: Mmm... pues en el posgrado, en el doctorado, a lo menos en Ciencias de la Tierra, hay lo que se llama Unidades de Investigación que debes cumplir al final de cada semestre; y eso se refiere a escribir un texto generalmente que, es un resumen de la investigación que hiciste durante el semestre. Y realmente la... son las únicas cosas que eh... escribí académicamente en español. Porque también el problema es que la ciencia se hace en inglés. Entonces, ya tengo artículos de publicaciones académica en revistas arbitradas, pero éstas son en inglés. Entonces, la mayor parte del tiempo, al final la escritura la hago en inglés y no español.

Investigador: ¿Y la tesis?

E67: La tesis, tienes la opción, de escribir una tesis completa todo en español, con mínimo un artículo aceptado en inglés, entonces; o escribirla por... gracias a tres artículos académicos. Entonces, en este caso la tesis está... ahh... en mayor parte escrita en inglés, sólo haces una pequeña introducción en español y luego todo lo demás es en inglés. Lo que creo que voy hacer, porque ya están a punto de salir los tres artículos; entonces, yo nada más voy a escribir una introducción en español y la conclusión.

Investigador: Y todo lo demás son los artículos que están en inglés, ¿pero podrían estar en español?

E67: Mmm... Si la revista es indexada puede ser en español; pero hay pocas revistas indexadas a nivel mundial, con factor de impacto bueno en español. Para el futuro de la academia, es mejor escribir en inglés y también para tener mejor

impacto en la comunidad científica mundial ¿no?

E69: [...] ensayos.

Sobre las dificultades que reportaron al momento de escribir en español, éstas fueron muy variadas: errores gramaticales (conjugación de verbos, artículos, plurales), uso de nexos y de conectores, de sintaxis o de puntuación. Sin embargo, destacaron que una de las mayores dificultades radicaba en conferirle un carácter académico a sus trabajos a través del léxico.

Categoría 13. Dificultades al escribir en español

E85: [...] todavía faltan muchos mmm... palabras, y todavía tengo muchos errores [no se entiende lo que dice], plurales, singulares.

E66: [...] pienso que el problema está un poco en la gramática en algunas estructuras gramaticales, pero sobre todo en... porque pienso que es jargón específico científico con que hay que usar en... los textos científicos. Entonces, esos conectores como construir las frases, la construcción de la frase es lo que me cuesta trabajo, para asegurar que sea científico, ¿no? Que no se ve como de vida cotidiana.

E78: Como los falsos amigos y algunas palabras pues tan parecidas que al momento de escribirlas que uno dice... no es posible que sea la misma.

E67: [...] al momento de escribir, lo que me falta realmente son las... como las palabras formales de la escritura. [...] Primero, tengo problema con los acentos, la ortografía. Bueno, es que también la ventaja es que en español, todas las letras que se pronuncian están escritas; lo que es diferente del francés, donde ahí a veces hay cinco letras en una palabra que ni siquiera pronuncian. Entonces, eso es fácil...

Investigador: ¿Pero no es algo que te preocupa demasiado?

E67: Bueno, Word luego me ayuda [refiriéndose a la corrección automática de la ortografía y acentos].

E69: No es muy académico, sí [refiriéndose a su español]. Bueno, ni en alemán me sale, ese es el problema. Entonces, hasta que en alemán me pasa lo mismo, estoy escribiendo y pienso: No, eso no se puede leer. [...] el vocabulario. Porque, pues quiero poner algo, pero pienso: Ah, no, eso es muy coloquial. Entonces, pues sí intento buscar otra palabra y... bueno, de la gramática lo que a mí me cuesta, pues es la cosa, que creo que cuesta a todos los alemanes, eso de indefinido e imperfecto, por ejemplo; pero eso, pues en los ensayos no se da mucho ¿no?

E76: Mmm... Pues me cuesta un poco mmm... Las conjugaciones, verbos eh... pero no es muy difícil. ... (piensa) gramática.

Sobre si lengua materna ayudaba o no a la escritura, al igual que con el grupo anterior, con lenguas afines concordaron que era una ayuda; mientras que con lenguas distantes, no.

Categoría 14. Influencia de la lengua materna para escribir en español

E85: A veces, como pronunciación; a veces muy similar [refiriéndose al Japonés y al español], pero para escribir, nada (risas).

E8: [...] como en la parte gramatical, creo que hay cosas muy parecidas entre checo y español y también la forma como construyes las frases, las oraciones, no es tan estricta como en inglés, es más libre como en español; así me conviene muchísimo.

E66: No, no tiene nada que ver, pienso que no hay ninguna relación [entre el polaco y el español]. No, para mí es como algo diferente. No lo considero ni como un obstáculo ni como una ayuda; algo diferente, para mí es tan diferente que, porque si conozco francés y eso, tal vez, es por el vocabulario si me ayuda más pero polaco no tanto.

E78: [...] italiano y español siempre son parecidos. [...] pues son lenguas latinas, seguro tengo más facilidad que un japonés. Yo creo que son lenguas latinas, tienen muchas cosas parecidas, la sintaxis es parecida, no igualita, eh.... La verdad es que nunca por mi lengua, bueno en el maya, pero lo que pasa es que quizás no tengas que aprender otra forma de pensar como, por ejemplo, si pasa con el chino o con el maya, sino que más bien es... no sé si una traducción, pero uno muy parecido entre el italiano, el francés y el portugués. Es muy similar. Sobre todo en cómo se piensa.

E70: Ah... bueno, en realidad no sé si es ayuda. Yo diría que puede ser hasta contraproducente, porque uno tiene ciertas costumbres cuando escribe un trabajo, digamos académico, en la lengua académica, y pues ah... a veces, como que se mete el concepto de como... organizar las ideas de... de... de... mi lengua materna [checo]. Así lo siento yo, y creo que puede ser más bien contraproducente que una ayuda en realidad.

E76: Creo... que no, porque la gramática es muy diferente. Ah... los dos idiomas, se diferencia mucho, pero sí que mmm... primero pienso ah... la idea que quiero expresar, en chino y luego... como traduzco, pero ahm... resulta que ahm... las frases están influenciadas por mi idioma y... no es tan bueno.

El proceso de escritura resultó muy variado en los estudiantes del grupo intermedio alto. Sólo una estudiante, quien puntuó con la nota más baja del grupo, reportó tener un proceso de traducción al español después de escribir su texto en su lengua materna. Los demás estudiantes manifestaron escribir sus textos en español. En algunos casos, mencionaron hacer uso del diccionario para escribir la palabra en español, mientras que en otros, utilizan su idioma materno intercalado con el español.

Categoría 15. Descripción del proceso de escritura (incluye la fase de lectura, planeación, uso de recursos materiales y de la edición del texto con ayuda de terceros)

E85: Ahorita, siempre yo escribo en japonés, mi lengua natal, y después yo... tra... tra... traduzco en español, por eso siempre yo utilizo diccionario, siempre, siempre, siempre. Pero está bien me gusta.

Investigador: Por ejemplo, cuando lees en español, ¿aun así escribes tu capítulo en japonés y después lo traduces?

E85: Sí (risas). Antes yo escribía en español desde el principio, pero me estaba confundiendo. Después, por eso eh... ahora, sí estoy escribiendo en japonés, y la cita nada más español original. [...] Pero nada más para mí, no voy a presentar mi escrito en japonés.

Investigador: En este proceso de traducción, ¿qué es lo que haces, qué recursos utilizas para ir escribiendo en español? O sea, ¿utilizas un diccionario, utilizas un traductor de internet?

E85: Primero, yo escribo con diccionario electrónico, español-japonés, y... después, si quiero saber sinónimo, etcétera, voy a... utilizo cambio de sinónimo, por Word en Word y (piensa)... A veces, elegir una otra palabra distinta.

Investigador: Y, por ejemplo, para [...] algunas frases más hechas, ¿cómo lo haces o cómo lo expresas?

E85: Ah... A veces, yo utilizo la palabra japonesa; en sí y después, yo añado la explicación, de esa palabra, como tal vez: *uki uki*, es serigrafía japonesa. Pero no hay palabra *uki*, lo que es letras cursivas bajo de texto explico sobre *uki uki* con unas frases.

Investigador: ¿Le pides ayuda a alguien para revisar tus textos?

E85: Ah... Sí (risas), mi amigo conoce muchas cosas en Japón. Me ayuda, a veces yo quiero decir algo en español, pero... sólo los japoneses saben... algunas culturas.

Investigador: ¿Lo haces con todo tu texto, o solamente con cosas especiales? ¿Alguna vez le has dado a tu compañero todo un texto?

E85: Algunas veces mi amigo lee todo, y dice: "esto es..." Y mi tutora también conoce que yo soy extranjera y todavía hay muchas dificultades, por eso me ayuda mucho para corregir cosas así.

E8: Yo primero, siempre pongo una estructura como la introducción, el contenido, la parte teórica, metodológica, y luego la conclusión y eso se me facilita mucho para ir avanzando y desarrollando mi texto.

Investigador: ¿Y ese caso lo has tenido en México? ¿Has tenido que escribir ensayos en español? ¿Fue el mismo proceso? ¿Hiciste lo mismo?

E8: Ajá. Yo creo que cuando uno lo tiene bien hecho, bien estructurado, en un idioma, pues no depende necesariamente del idioma. Entonces, si uno lo tiene ya como.... internalizado. [...]

Investigador: Pero, en ese momento en que lo que te importa es transmitir el contenido ¿hasta qué punto el pensar en la ortografía, por ejemplo, tú empiezas a ser consciente de esas cosas? ¿Tu atención se fija un poco en la sintaxis, en cómo es la oración, en las cuestiones formales?

E8: Creo que sí. También es una parte necesaria; luego cuando voy revisando los textos, dos, tres veces me fijo más. Sí, pero creo que es como necesario, que no son como dos cosas separadas, que no es contenido de otro, sino parte de la sintaxis y no sé, la morfología. Eh... Que son cosas que van a mano. [...] al principio, cuando llegué a México, tuve pues... que escribir un texto en español; empecé en inglés y luego traduje a español. Y pues ahora no. Ahorita también como que me ayudaron los textos, las lecturas, la cantidad de palabras, los términos como utilizados... como los diría, como la... terminología.

Investigador: Y ahora, recientemente en los últimos trabajos ¿Qué haces? O sea, empiezas a escribir en español, pero cómo es ese proceso; o sea ¿todo es en español o metes palabras...?

E8: Ya es como todo; es como que ya no necesito... a veces también me viene como esta... esta impresión o este sentimiento que... cómo lo haría o cómo lo escribiría en inglés. Pero pues ya no intento porque es mejor, como después de estos cinco meses aquí... y meterme como en este reto, y...

Investigador: Pero, ¿qué haces, por ejemplo, en ese momento donde te vienen estas palabras que te pueden venir en checo, que te pueden venir en inglés? [...]

E8: Exacto, así es. Y como también lo busco en el diccionario, como la palabra correcta y luego como... cuando quiero como colocar las palabras y pues no estoy seguro si la frase está bien y entonces...(risas) lo intento como de una manera más sencilla, sí.

Investigador: ¿Eso interfiere en tu proceso o...?

E8: Sí, a veces sí. Es como... ¿Qué quería decir? ¡Sí claro! si no hubiera este proceso, este momento de buscar las palabras de...ah... en el diccionario; entonces, sería más fluido el pensamiento. [...] Lo que me pasa es que... lo que pasa con los textos, ese es mi problema, es como mi manera de escribir, igual tanto en checo, en inglés como en español es que... yo de repente en el mismo momento quiero como tener la frase ya hecha. Entonces, por eso me tardo muchísimo a veces. Ya pues... ya intenté a cambiar como este estilo de escribir porque es muy inefectivo... es como que a veces te tardas muchísimo, entonces... luego pues, revisas la parte y buscas las formas correctas, y las palabras finas, no sé qué.

Investigador: ¿Le pides ayuda a alguien?

E8: No, (risas). No, justamente siempre lo mandé a mi profesor, a mi profesora.

E66: En general pregunto primero si tengo que entregar el texto en español o inglés, porque de vez en cuando puedo entregarlo en inglés si es para una conferencia, por ejemplo, en otros países, sí me permiten hacerlo en inglés. Entonces, ese es el primer filtro. [...] Cuando no tengo opción entonces escribo en español y... en general intento escribir desde principio en español. Entonces, no es que escribo en inglés y después traduzco, es desde principio en español y en general mi forma de escribir el texto es: primero me pongo las preguntas y el esquema general del texto ¿Qué quiero decir? ¿Cuáles hipótesis? de forma más estructurada ¿no? para organizar. [...] Hago... mi esquema y también preguntas para ver, porque tengo mi esquema y después estoy viendo, ok me faltan esta... me falta esta información, esta información, entonces lo busco y cuando tengo que buscar en internet, sobre todo en algunos libros y si tengo que escribir un texto en español voy a buscar más páginas en español, para ya tener esta referencia en español, preferiblemente ¿no? Porque si no encuentro datos en español, voy a buscar en otros idiomas, pero... cuando tengo que escribir en español preferiblemente busco en español.

Investigador: ¿Y cuándo vas? O sea me imagino que... bueno, el proceso de ir escribiendo en una segunda lengua es, de pronto no encuentras las palabras o...

E66: No, es... pienso, primero tengo una frase en la cabeza puede ser que está formulada en inglés o polaco, algo que me falta, inmediatamente busco, porque quiero escribir inmediatamente la frase.

Investigador: O sea, todo es en español.

E66: Todo es en español... solamente si no tengo conexión al internet o no tengo un diccionario a la mano. Entonces, sí lo pongo en otro idioma o de vez en cuando si no estoy segura de la forma gramatical, cómo ponerlo en español, voy a poner dos versiones: la versión español y al lado lo pongo en inglés, por ejemplo, o polaco; la palabra o la frase para estar segura que aquí está lo que yo quiero decir.

Investigador: ¿Le pides ayuda a alguien?

E66: [...] en general intento escribir con mucho tiempo de anticipación para tener este tiempo y ver con mis amigos, con otros alumnos, les pido ayuda para... para verificar mis textos.

E78: [...] en el mío proceso de escritura, pues, sí claro, comienzo más como en italiano y traducir. Pero como al principio y ahorita es en español, pero algunas palabras son en italiano, más bien quizás es que como la diferencia etimológica de un término a veces se me escapa en extranjero. Términos que son sinónimos, pero no tanto, pero que tienen diferentes usos y entonces vas buscando el que más se acerca a lo que quieres expresar.

Investigador: ¿Le pides ayuda a alguien?

E78: Bueno, las tareas y ensayos que entregué, claro, los hago revisar con un amigo, amiga. [...] También es difícil preguntar a un amigo o amiga: oye, tengo un texto de 50 cuartillas. Entonces, pues sí, a veces pasa. Si son 10 o 15 páginas, quizás sí.

E67: Siempre mi manera, de construir un texto, en español; pero también en inglés, es escribir como pienso todas las ideas. Una vez que tengo un texto largo, voy a empezar desde cero y empezar a estructurar todo: esta frase la puedo recortar aquí, no está tan clara; esto no tiene sentido. Y una vez que el texto tiene una forma, una arquitectura mejor; como es en español, sí voy a tener que pasar con un amigo que me relea todo y que me dé su visto bueno, ¿no? [...] Siempre estoy en el diccionario Word reference; y a veces, me viene la palabra en inglés; escribo tanto en inglés. Entonces, a veces la busco del inglés al español; a veces del español a... Y a veces, me ha pasado de saber una palabra en español y yo buscarla en inglés. [...] Word luego me ayuda [refiriéndose a la ortografía y acentos].

Investigador: ¿Le pides ayuda a alguien?

E67: [...] sí voy a tener que pasar con un amigo que me relea todo y que me dé su visto bueno.

E69: Mmm... No, bueno, por ejemplo, si yo escribo un ensayo, pues leo lecturas, ¿no? Entonces, creo que es lo que... lo que sí tarda un poco más es... formar las ideas en español; porque los textos normalmente están en inglés, por ejemplo. Entonces, eso sí; porque a veces no me sale la palabra alemana y tengo problemas de [no se entiende muy bien, parece que dice traducción] y no sé qué; entonces, en alemán como es lengua materna tengo la lectura y sí, sé cómo escribirlo, ¿no? Pero en español, eso creo que sí toma más.

Investigador: Cuando tú lees en español y tienes que escribir en español, y cuando tú lees en inglés y tienes que escribir en español, ¿qué dificultades ves? ¿Es igual de fácil escribir?

E69: Pues, a mí me gusta más... este..., leer en inglés y escribir en español, porque no tienes tanta... como que no es tan fácil copiar, ¿no? Copiar así, porque yo no quiero tradu... Porque si ya está en inglés, y además así, en un inglés muy bonito, pues no lo voy a copiar, porque eso no se hace; pero como que... sí, es mucho más difícil así, decirlo en otras palabras, si ya las palabras están usadas, ¿no?

Investigador: ¿De qué recursos te vales... o cómo le haces tú para ir construyendo tu texto?

E69: Pues normalmente, sólo pues un diccionario, uso leo, creo que sólo hay en alemán online. Porque los diccionarios [se refiere a diccionarios en papel], no, me tardo casi un mes buscando la palabra (risas). Entonces, ahí, porque ahí también tienen este... También ponen, por ejemplo: Eso sólo se dice en España, en otro país, o así, dichos. Y pues también así... Ah, y también uso *WordReference*, porque ahí también hay como sinónimos, antónimos y eso, pero ya.

A la hora de ir construyendo tu texto, ¿cómo te vienen las ideas?

E69: No. Bueno, yo leo, por ejemplo, leo una lectura o algo, ¿no? Y pues, sí ya me apunto las ideas y los voy acumulando, y este... pero ya todo en español. Entonces, ya como que veo lo que tengo y lo pongo y pongo mis ideas, pero ya todo...

Investigador: ¿Y si esas palabras no te vienen en español?

E69: Los busco enseguida, ajá. Cuando me los apunto, es cuando ya enseguida los busco. En eso, normalmente, en el proceso, eso me tarda más bien como... De todo, como que... poner

todas las ideas así, en la compu y pues ya en español. Como fichas, eso lo hago, pero en la compu.

Investigador: ¿Le pides ayuda a alguien?

E69: Me da pena (risas). Este... bueno... para, por ejemplo, para mis cartas de motivación, cosas así, sí se los pido a mis amigos, pero... así para un trabajo escrito... Es que no me parece tan... como que no es una tesis, ¿no? Entonces, no me parece tan importante para que lo tengan que leer, porque normalmente son como veinte páginas y, si a mí alguien me da algo para leerlo, yo siempre estoy así de: Ay, bueno, pues sí lo hago pero... sí es muy pesado.

E70: Bueno, pues ah... (piensa). Estoy tratando de pensar desde el principio en español. Claro, a veces es inevitable que sí me falta una palabra y ahí pues tengo a mano abierto el traductor y estoy buscando la... la palabra correcta, sí. Pero la base la hago en español directamente.

Investigador: ¿Y te ha pasado que hay alguna idea que no sabes bien como formularla en español y la escribes en checo?

E70: No, digamos que más me pasa eso que me falta la palabra. Que sé qué quiero decir, pero no encuentro la palabra, pero no que toda la oración me viene en checo y... Yo tengo que traducir todo eso, no.

Investigador: Y cuando tienes idea, por ejemplo, ¿cómo se dice esto en español? ¿A quién le preguntas o qué haces?

E70: Bueno, lo busco normalmente en internet. [...] Pero claro, ayuda también el Word, que tiene la corrección ortográfica, y luego uno puede buscar en internet o... En algún manual gramático para... resolver tus dudas. [...] todo el tiempo escribo en el Word; así que tengo la corrección y este tipo de cosas.

Investigador: ¿Le pides ayuda a alguien?

E70: Necesito que alguien lo cheque, porque no podría entregarlo tal cual (risas).

E76: Primero pienso ah... la idea que quiero expresar, en chino y luego... como traduzco, pero ahm... resulta que ahm... las frases están influenciadas por mi idioma y... no es tan bueno. [...] Normalmente, no voy a escribirlo en chino porque bueno, no tengo mucho tiempo para ah... Sí, que pienso un poco en chino, no directamente en español.

Investigador: Pero, haces el esfuerzo de escribir siempre en español.

E76: Mmm... sí, si tengo que escribir algunas cosas de español pienso en chino y en español.

Investigador: Me imagino que al momento de escribir, vienen ideas en chino, palabras en chino y entonces, si no la conoces en español, ¿qué haces?

E76: Pues busco otras palabras, o intento expresarlo de otra forma.

Investigador: ¿A veces escribes un poco en chino mientras estas escribiendo en español?

E76: No, porque son muy diferentes (risas).

Investigador: ¿Le pides ayuda a alguien cuando escribes un documento? Le dices a algún compañero: ¿puedes revisarlo? ¿Qué tipo de ayuda pides?

E76: Mmm... Normalmente voy a consultar como diccionarios, o investigar por internet.

En cuanto a la retroalimentación que reciben de la lengua, generalmente ésta fue amplia por parte de sus tutores, cuando existían errores sustanciales de estructura; sin embargo, en los demás casos, cuando no se trataba de los tutores, ésta era mínima o nula.

Categoría 18. Retroalimentación que recibe sobre su escritura en español

E85: Es depende de los maestros. Mi tutora siempre cor... cori... [Le cuesta trabajo]... Corrige las palabras muy detalladas; y en doctorado tengo un tutor principal y dos tutoras; y otro tutor siempre corrige mi idea; y otros casi no, sólo dicen: está bien o no.

Investigador: ¿Esa retroalimentación es suficiente?

E85: Mmm... No, no quiero, quiero hacer más (risas). Es que... Por mi tutora, corrige las palabras pero la idea casi no, como no me digo sinónimos, etcétera; sólo si hay error corrige, por eso quiero saber más sinónimos o más palabras distintas, etcétera; para mmm... Profundizar, enriquecer mi tesis.

E8: Nada.

Investigador: ¿Nunca hay como comentarios?

E8: Desgraciadamente no. Hay comentarios del contenido, pero nada de...[lengua]

Investigador: Pero no hay como: ten cuidado de tus acentos...

E8: No.

Investigador: Revisa la puntuación... ¿No? Nada, nada de lo que involucra la escritura.

E8: No, nada. No (risas). Es por eso que estoy muy agradecido con este curso [refiriéndose al taller de escritura que se les dio por haber participado en el EXELEAA].

E66: Contenido casi no, contenido en general les gusta, es la forma gramatical, es...

Investigador: ¿Te lo corrigen ellos? ¿Te corrigen ahí?

E66: Sí, sobre todo mi tutora personal; porque con ella tomé esta clase de... teorías de comunicación. Entonces, para este seminario de teorías de comunicación, sí, ella me... corrigió este texto y dice: no, sería mejor usar esta frase, o esta frase sería más académico si lo pones de esta forma.

E78: pues depende cuánto tiempo se toman, ¿no? Porque, por ejemplo, la corrección de los acentos puede ser muy rápido. Y... luego la corrección del contenido a nivel de contenido en sí o de expresión del contenido, pues... depende. Algunas veces, o sea, no me acuerdo, pero alguna vez algunas ideas no fueron entendidas a causa de la lengua. Pero casi nunca. Una o dos veces.

E67: De la parte del comité, no mucho.

Investigador: Tú tutor principal, ¿sí lo hace?

E67: Sí [...] También, es que mi tutor es francés. Ya hace quince años que vivió en Chile, en Ecuador, ahora en México. Desde quince años que está en un ambiente hispanohablante. Entonces, sí habla muy bien, pero no perfectamente. Entonces, sí me corrige algunas cosas, pero... no todo.

E69: Pues en el taller que hice, por ejemplo, sí estamos así, íbamos avanzando, ¿no? Entonces, yo cada vez le enviaba: Mira, mi título es así. Y lo discutíamos: Mi objetivo así. Entonces, ahí se lo mandábamos por e-mail. Entonces, ella sí nos dijo: Ah, mira esto lo escribiría yo así o... añade también esa idea. Entonces, ahí sí hubo mucho feedback y... pues, al final me dijo que todo muy bien pero... Sí.

Investigador: Pero entonces, esta retroalimentación era no solamente sobre el contenido, sino también un poco sobre las ideas.

E69: Sí, como sobre las ideas y pues también sobre el español, mucho; pero ya vi que con el tiempo, siempre me corregía así, menos, ¿no?

Finalmente, cuando se les preguntó sobre qué consideraban que podría mejorar su escritura académica, al igual que en el grupo intermedio bajo, se enfatizó que la práctica de escribir aunado a la lectura y la retroalimentación de trabajos escritos ayudarían a mejorar la escritura.

Categoría 19. Reflexión sobre qué podría mejorar su escritura académica

E8: Ahorita también como que me ayudaron los textos, las lecturas, la cantidad de palabras, los términos como utilizados... [...] Entonces, eso me facilitó mucho para expresarme.

Investigador: ¿Los textos que leías estaban en inglés o en español?

E8: Pues en español, porque nos dieron montón de textos, lecturas, que... para cada clase, pues teníamos que leer los textos y... pues estar preparados, y eso me ayudó mucho para pues ampliar mi vocabulario y darme la confianza a... y ver los verbos, las palabras, los nexos y luego utilizarlos en mi mismo texto, en mi mismo trabajo.

E66: A mí me gustaría tomar algunas clases de gramática como nivel más avanzado, como condicionales, subjuntivo, tomar estas clases de gramática y escritura, aunque no me gusta, pero sí, yo sé que lo necesito y no sé...

E67: Yo creo que, más bien leer, eso es. Cuando leo un texto siempre me da gusto ver cómo está formada la frase, y decir: ¡Ah, sí! Así se dice. Y cosas así, leer ayuda mucho; pero el problema es que en la academia, como ya lo he platicado, no se lee mucho en español. Pero eso realmente, después de haber hablado muchísimo, ahora lo que me falta es leer en español.

E69: Mmm... Pues leer más en español (risas). Eso sí, mmm... Pues no sé, igual sí escribir más en español, pero que también me corrijan [...] Creo que, pues escribiendo ya se aprende, ¿no? Pero sí, se necesita cierto *Feedback* de los profesores o de... quién sea.

E70: Bueno, para elevar mi escritura, creo que una cosa sería, ah... (piensa), ver más en la cuestión de la gramática; sobre todo de la puntuación, también para ayudar la lectura. Eso es lo fundamental, creo yo y... No sé, bueno, estaba pensando que tal vez podría tomar un curso de redacción que ofrecen también en el instituto, pero... todavía no me he decidido, depende de otras asignaturas y del tiempo.

E76: Mmm... que como... por ejemplo, unos libros, ahm... de unos autores o profesores, que son ahm... muy bien escritas que puedo leerlo, investigar, creo que sí (risas).

6.4.2.3. GRUPO AVANZADO

El grupo de nivel avanzado fue el más numeroso con 12 estudiantes: siete de maestría y cinco de doctorado. Los estudiantes del nivel avanzado reportaron un tiempo en México entre 3 meses y 6 años. En cuanto al tiempo de escribir en un contexto académico en español, el rango variaba entre 4 meses y 8 años.

Categoría 1. Tiempo de vivir en México**E79:** Dos años**E86:** Como dos años y medio**E80:** Tres meses**E68:** Cuatro años y medio**E83:** Seis años**E74:** Casi cuatro meses**Categoría 2. Tiempo de escribir en un contexto académico en español****E79:** Dos años**E84:** Cuatro meses**E86:** Cuatro meses**E80:** Como cinco años**E4:** Hace dos años**E83:** [8 años]

La autopercepción del nivel de escritura en español fue muy diversa. Hubo quienes afirmaron que su nivel era principiante, más bajo que en su lengua materna, intermedio, bueno, bastante bueno y avanzado. Quienes afirmaron que su nivel era avanzado, matizaron que éste podía mejorar. Con respecto a las sensaciones que reportaron al momento de escribir, fueron todas positivas. Sólo una estudiante mencionó que se sentía chistosa al escribir en español, pues sentía que utilizaba estructuras muy enredadas. Los estudiantes del grupo avanzado sentían que podían expresar bien sus ideas, tenían confianza y escribían con placer y comodidad en español. Sin embargo, algunos también manifestaron que esta confianza no siempre era positiva, pues podía conducir a errores. De igual forma, dos estudiantes manifestaron que el placer por escribir en español no se comparaba aún con el gusto por escribir en lengua materna.

Categoría 3. Nivel de escritura en español**E20:** Avanzado**E84:** Eh... básico, este... nivel principiante**E86:** Mal**Investigador:** ¿Mal?**E86:** Bueno no, la verdad no lo considero tan mal, pero en comparación con mi nivel de... escritura en inglés pues es mucho más bajo, pero cada vez se ve como a lo largo del semestre, se ha mejorado y eso pues, es muy positivo ¿no?**E80:** Intermedio, mmm... avanzado, más o menos, todavía hay parte que necesito**Categoría 4. Sensaciones al momento de escribir en español****E20:** Siento que puedo expresar bien mis ideas y expresar mis conceptos y que tengo más recursos de vocabulario que antes. [...] Me siento con bastante confianza.**E15:** [...] no tomo tanto placer como escribir en francés. Cuando escribo en francés sí, me gusta mucho, pero en español todavía no, pero... por ahí va, no sé si se puede llegar un día...**E84:** Absolutamente con confianza, pero eso eh... no creo que sea totalmente positivo porque eso me lleva ah... hacer muchos errores

perfeccionar y a veces cometo errores gramáticos que no me di cuenta...

E82: Eh... yo considero que bien (risas). Sí cumplo como los requisitos.

E74: Creo bastante Bueno.

E80: Tengo confianza en mi ah..... redacción de trabajo. Sin embargo, siempre y cuando termine un trabajo necesito ah... dar un vistazo o revisando varias veces para asegurar que no cometiera algún error gramático

E4: Es una frustración de no tener la misma habilidad. [...]Escribir es algo que me gusta, que me cuesta trabajo, que no me sale necesariamente tan fácilmente pero es algo que disfruto y sí lo veo realmente como una aventura.

E72: [...] me siento bien cómoda, con... mi manejo de... del español escrito, pero como ya dije, también depende ¿no? [refiriéndose al tema].

Todos los estudiantes avanzados coincidieron en que el tiempo que les tomaba escribir en español era muy similar al que les tomaría escribir en sus lenguas maternas o en inglés, aunque también matizaron que dependía del tipo de texto que escribían.

Categoría 5. Tiempo que toma escribir en español

E79: [...] un poco más, pero ya no tanto.

E20: Creo que es un poco más. Depende que estoy haciendo. Si estoy relacionando cosas que escribieron en inglés y español, necesito más tiempo, me cuesta más. Si solo estoy escribiendo mi opinión, lo que yo entendí, siento que es más fácil. Me tardo más cuando tengo que integrar el material, sobre todo cuando está en inglés. Porque en mi cabeza tengo que traducirlo primero. Ahí creo que no resumo, solamente busco la traducción. Y me tardo mucho por términos o porque a veces no sé cómo se dice. A veces leo la palabra en inglés, pero no me acuerdo cómo se dice en español; entonces, tengo que pensar o buscarlo en un traductor.

E84: Mmm... poquito más, pero, más o menos el mismo.

E80: [...] no creo que cuesta más tiempo [...] cuesta como mismo tiempo que en chino...

E4: [...] se va reduciendo la diferencia, creo. Porque en francés también hay una reflexión, hay una estructura, hay una versión final, hay palabras que buscar en el diccionario, hay una revisión que hacer en francés también. Entonces, sí me cuesta más, pero no mucho más y se va reduciendo con el tiempo.

E72: [...] Sí se requiere más tiempo para escribir en español [...]uno tiene que pensar un poco más y reflejar [quiso decir reflexionar] un poco más, porque las ideas, sí fluyen pero para... para traducirlo o quizás no traducir, porque uno no hace traducción, pero sí interpretarlo en un

idioma que no es el suyo y que... que haga sentido, también, bueno sí, eso... eso, bueno demoro mucho más en... en escribir en español.

En relación con la categoría cursos de lengua, los entrevistados mencionaron que una clase de español era innecesaria. Comentaron que tendría que ser un curso muy específico o dirigido propiamente a la revisión individual de sus textos.

Categoría 6. Toma clases de español o manifiesta una opinión al respecto

E20: [...] cuando entré a la maestría, yo sentía que si tomaba un curso aparte, que iba a ser un trabajo extra. Entonces, sentí que no era lo que buscaba y bueno... yo pensé: “mi español no está tan mal” Y como los profesores no me dijeron nada, bueno pues pensé que no necesitaba clases. Personalmente, pienso que no necesito un curso completo. Especialmente ahora, pienso que la retroalimentación que necesito es del trabajo que estoy haciendo. Tendría que ser algo muy específico, como lo que hizo mi tutor en la maestría. Sé que hay cursos de escritura en otras facultades, pero con todo lo que estoy haciendo en el doctorado, pues no sé si lo necesito.

E84: No creo que me sirviera un... un curso, por la modalidad de aprendizaje que tengo yo; o sea, que no... no consigo enfocarme mucho cuando tomo cursos de estas cosas.

E82: [...] lo que podría ayudar sería como... como un curso, pero la verdad es que no lo he buscado (risas). Sí, sí podría tomar un curso donde explican con cosas muy específicas y como eh... como variar, por ejemplo, ciertas expresiones para no... no estar escribiendo cada rato: sin embargo, no obstante, sin embargo, no obstante. Para que no se repita. [...] yo considero para mí mmm... que ya no necesito un curso de eh... de redacción para extranjeros, sino tal vez en general, para ver dificultades que muchos tienen...

Los estudiantes reportaron que habían tomado o estaban tomando varias materias en el posgrado. En algunas ocasiones, estas materias eran cursadas como oyentes o como materias no curriculares para el doctorado.

Categoría 7. Clases en el posgrado

E15: Yo voy a dos clases porque yo pedí, para tener un ritmo, para encontrar a profesores con una relación más... los ves cada semana. Y también he encontrado gente interesante. Entonces, no tengo obligación, pero voy a dos seminarios este semestre.

E86: En este semestre cursé cinco materias.

E80: Ah... yo este semestre llevo tres materias ah... son inscritas y otras ah... dos materias allá como oyente.

E68: Eh... no está previsto, este seminario, eh... este semestre, no había seminario de investigación pero... sí quiero estar en un seminario de investigación con mi directora de tesis el próximo semestre y quiero ir como oyente libre a materias de la licenciatura.

E72: Bueno, ya no tenemos que cursar materias. [...] Sí, el primer año nos tocó... nos tocó materias, porque el primer año del posgrado del doctorado es escolarizado. Entonces, sí nos tocó ir a asistir a clases; así como 2 clases por semestre, 2 materias, por semestres. Yo hice como 4 cada semestre.

E74: Cuatro

En cuanto a las lecturas asignadas, éstas eran en su mayoría en español, aunque ocasionalmente había material de lectura en inglés. En cuanto a las lecturas que ellos elegían, el idioma muchas veces se encontraba dado por el tema de la investigación, aunque también para algunos era más fácil leer en otro idioma porque era más claro.

Categoría 8. Idioma de las lecturas asignadas

E79: 80% en español... pero son muchos tro... y las... las este 20%, lo que queda de 20% en inglés.

E15: En español, mucho, lo que también sí ayuda mucho, pues tienes la información en español, luego escribirla en español es más fácil. Entonces, español, luego en francés y en inglés, pero más que todo en español y en francés.

E84: En español, en inglés y en francés. [...] Mayoritariamente español, sí.

E80: Principalmente el español y tampu... también eh... un curso tenemos frecuentemente lectura en inglés.

E72: Están en español

E74: Siempre en español

Categoría 9. Idioma de lecturas adicionales

E79: [...] yo a mi propia cuenta ya estoy leyendo mucho más en inglés.

Investigador: ¿Por qué?

E79: Porque es más fácil (risas).

E20: Pues he leído mucho de aquí porque mi estudio es de aquí, pero uno 80% en inglés, un 15% en español y un 5% en alemán, porque como no hay mucho material de aquí, tengo que leer en inglés.

E84: No, principalmente en inglés, pero yo creo que esto es... tenga mucho que ver con lo que estudio, Ciencia Política y mmm... mucho material es en inglés.

E4 (estudiante quebequense): Si puedo escoger entre algo escrito en inglés y en español, lo escojo en inglés; y si es entre un francés de Francia, prefiero en inglés, me gusta mucho la manera en que escriben los ingleses en general.

E82: En español, inglés, alemán, francés, eh... [...] es como... 20 o 30 por ciento de español y entonces serían como... 60, 70 por ciento inglés y un poquito de alemán. [...] para mi tema estoy trabajando con más textos en inglés.

E74: Si tenía, por ejemplo, presentaciones en español pues las buscaba en español porque es más fácil o, por ejemplo, he tenido unas

presentaciones sobre Italia: la democracia en los partidos políticos, y en este caso los busqué en italiano porque había más material.

Sobre el idioma en que toman sus notas de clase, en general, mencionaron escribir todo en español, aunque algunos estudiantes reportaron que de vez en cuando introducían algunas palabras o notas en otra lengua, esto con el fin de recordar conceptos o ideas importantes.

Categoría 10. Idioma de notas de clase

E79: ¡Ah! No, están en “espanglish” total, todo, ya todo está como... A veces anoto cosas, estoy trabajando con gente allá de lo que sea y escribo en español y... no, está súper revuelto...

E20: En español, siempre. Eh... en general en español. Lo que me pasa mucho es que ciertas palabras sé la palabra en inglés, no me acuerdo o no sé la palabra en español si el texto está inglés, mis notas están en inglés y luego tengo que traducirla. Si estoy haciendo notas y me llega la palabra en inglés, veo ya en el texto cómo la puedo parafrasear.

E84: Todo en español, sí porque, se me hace más fácil como... de escuchar y escribir nada más lo que escucho ¿no? Porque... también es más fácil ¿no? Cuando llegas en un país extranjero empezar a pensar en la misma lengua española y entonces, en el momento que estoy en una clase y ya tengo personas que eh... en la misma clase que hablan solamente español, entonces ya me... me olvido de ser italiana, me olvido de hablar italiano o cualquier otra lengua que hablo y ya me... me enfoco en... solamente en español.

E86: Ahm... Ya siempre en español, siempre en español. A veces, si la lectura está en inglés, hago mis notas en inglés; por siempre refleja el idioma de la lectura ¿no? [...] Pero, en general escribo en español, siempre. Hasta que a veces las lecturas están en inglés y escribo mis notas en español, porque voy a tener que hablar sobre ellas en español.

E4: [...] Si es un concepto que entiendo bien, lo resumo en francés. Si tengo la idea clara de lo que significa, voy mucho más rápido tomando notas en francés.

E80: Ah... la mayor parte, ah... la mayoría de mis notas apuntes lo hago en español y cuando aparece palabra con que no conozco la voy a buscar por diccionario y luego voy a hacer apuntes en chino para recordar.

En relación con los exámenes escritos, la gran mayoría del grupo reportó que no había exámenes escritos en clase o muy pocos. En contraste, los estudiantes mencionaron diversos trabajos escritos que les son solicitados para evaluar su desempeño académico (ensayos, trabajos académicos o capítulos de tesis).

Categoría 11. Experiencia con exámenes escritos

E79: No... yo creo que no. No, ninguno.

E15: No, nada. Nada, nada, nada.

E84: Sí, nada más en una mmm... en una clase.

E80: Ah... ya no tengo exámenes.

E68: Mmm... en mi carrera hay muy pocos exámenes. [...] En el posgrado no, no me ha tocado.

E74: Exámenes escritos en clases, no.

Categoría 12. Experiencia con trabajos escritos

E79: es que como no mmm... no me siento que he estado escribiendo tanto, tanto, tanto como en otras disciplinas, en otras disciplinas [...] voy entregar ensayos... rigurosos, académicos... y bueno luego como cada semestre hay que entregar uno también, luego hay todos los pasos demás de la tesis, que no es cualquier cosa.

E15: Sí, uno es trabajo en casa y tengo que entregarlo al final del semestre y otro también trabajo en casa e hice presentación en clase. [...] El trabajo es presentar los avances de mi tesis y vinculándolos con lo que vimos en clase, en qué medida lo que vimos los textos, la teoría a mí puede ayudar para entender mejor mi objeto de investigación.

E80: [...] hace dos años elaboré mi primer trabajo largo como la tesina a nivel ah... licenciatura. Fue el trabajo más largo que había elaborado en aquel entonces. Ya tengo más confianza en la redacción de trabajos largos [quiso decir largos] como informe o avances, este... de proyecto que estoy elaborando. [...] necesito entregar trabajos ah... para cada materia a terminar el semestre todavía quedan pendiente todos trabajos que entregar.

E83: [...] Aquí en la facultad, en estos últimos como 3 o 4 meses sí he escrito bastante más en español que antes, pero 4 meses no va perfeccionarlo tanto.

E74: [...] desde el dos mil once [cuatro años], porque tenía muchos ensayos muchos trabajos para entregar.

Sobre sus dificultades al escribir en español, éstas son de distinto tipo, aunque de manera general, todos reportaron una falta de léxico lo suficientemente sofisticado o técnico para conferirle un carácter académico a su discurso escrito. Por una parte, en cuanto al nivel oracional, reportaron diversas dificultades de puntuación, sintácticas, ortográficas, gramaticales (género, conjugación de verbos, preposiciones) y léxicas. En los aspectos macro estructurales, la formación en lengua materna ayudaba, aunque se apreció una dificultad para escribir de acuerdo con las convenciones académicas apropiadas.

Categoría 13. Dificultades al escribir en español

E79: [...] encuentro problemas de este... este sin... sintácti... sintácticos. Este... sí, complicación de ideas, problemas con cláusulas, a veces aquí y allá como... confuso femenino y masculino. O sea, errores súper típicos que todos tenemos ¿no?... [...] O conjugaciones. [...] otra exigencia este... el ambiente académico.

E15: [...] La sintaxis, eh... también, sí creo, pero más, yo diría, más el vocabulario. Me siento súper limitada en eso. La sintaxis también, creo que lo tengo con la sintaxis es que muchas veces puede estar parecida al francés, sin embargo, también hay construcción muy española, y eso, pues poco a poco siento que la voy agarrando, pero todavía no es un automatismo. [...] eso, de hilar las ideas, no tengo problemas, pero igual que con el vocabulario uso siempre las mismas palabras. [...] En Francia, desde la prepa, te aprenden (quiso decir enseñan) una metodología muy rigurosa. Sí, para eso estamos muy formados yo creo, la problemática, estructurar el texto, sus partes, entonces lo fui adquiriendo.

E20: Del inglés, me ha ayudado la estructura del texto. Esa parte la tenía muy fuerte y sigo escribiendo igual en español. Yo sé que en alemán, en muchas cosas, la estructura es diferente, pues a veces las ideas principales van al final de párrafo, mientras que en inglés es muy claro al inicio del párrafo.

E84: [El italiano] es un poquito más flexible, sí. Entonces... no puedo simplemente traducir del italiano, sino tengo que aprender un lenguaje específico del español escrito.

E4: [...] Es que comparo mucho con mi lengua materna o con el inglés, de repente, y siento que: ¿eso lo podría decir tan bien en francés! Es una frustración de no tener la misma habilidad. Sino, pues la sintaxis, que sí de repente se puede ver que tengo muchos galicismos, ¿no? [...] En español, de repente, el estilo, no sé si es el estilo mexicano se me hace muy enredado.

E83: [...] lo que me cuesta más, es quitar los anglicismos que tengo ah... O sea, no necesariamente pongo palabras que significa algo diferente, lo que hago es intentar expresarme en una manera que no sería natural por un mexicano o colombiano o español, o alguien... [...] el léxico, la manera de escribir; estoy pensando todavía un poco como que estuve mmm... como traduciendo desde inglés a español, pero no tiene tanto sentido en español decirlo como así.

Sobre si la lengua materna ayudaba a escribir en español, aunque es cierto que reconocieron que las lenguas romances y con cierta afinidad ayudaban al flujo y proceso de escritura, también advirtieron que esto no siempre era beneficioso, pues resultaba difícil hasta cierto punto discernir cuál era la estructura del español y cuál era de la lengua materna.

Categoría 14. Influencia de la lengua materna para escribir en español

E79: Bueno, en comparación con el mandarín ¿no? Pues obviamente sí, porque se relacionan ¿no? Mmm... A veces no, porque [...] nos apoyamos tanto [que...] el inglés como puede complicar un poco creo, la expresión a alto nivel [en español].

E20: [...] Del inglés, me ha ayudado la estructura del texto. Esa parte la tenía muy fuerte y sigo escribiendo igual en español. Yo sé que en alemán, en muchas cosas, la estructura es diferente, pues a veces las ideas principales van al final de párrafo, mientras que en inglés es muy claro al inicio del párrafo. Nunca me han dicho que la estructura está rara. No me acuerdo bien en la maestría, pero creo que sólo vi que la estructura de escribir en español era muy similar. Yo apliqué lo que sabía del inglés y vi que no había mucha diferencia.

E15 (estudiante francófona): [...] la ventaja es que también, aunque no sé si es, es que son idiomas, bueno, romanas. Creo que sí ayuda mucho. O sea tampoco tengo que cambiar toda la estructura gramatical.

E84: [...] Como hay muchas palabras que se parecen con el italiano, y esto me ayuda en la escritura también. Pero he... sí tienes que considerar que la escritura en español tienes unas estructuras diferentes y a veces eso no me ayuda completamente porque me confundo ¿no? Entre mi idioma y el español y entonces, a veces es una ventaja y otras veces no. [...] no puedo simplemente traducir del italiano, sino tengo que aprender un lenguaje específico del español escrito.

E80: [...] Sí, es otra sistema, eh..., mmm de habla, mmm la lengua entre el chino y el español es tan distinto que no puedo acudir a mi lengua materna para ayudarme eh... en el aprendizaje del español, por eso [quiso decir por eso] los otros chinos ponemos mucha énfasis en la aprendizaje de gramática y esta parte cuesta mucho trabajo.

En relación con el proceso de escritura, los estudiantes reportaron en general un excelente dominio de la escritura académica en español. Evidenciaron que el proceso de escritura reflejaba ya un pensamiento y procesamiento de la información en español. De igual forma, algunos de ellos enfatizaron que previo al proceso de escritura había un periodo de lectura con formulación de conceptos e ideas de cómo planificar la escritura incluso desde la lectura. La gama de estrategias que desplegaron para escribir fue variada; es decir, dependiendo de su conveniencia, no siempre escribían todo en español, sino que en ocasiones se apoyaban de su lengua materna para escribir con mayor rapidez y darle fluidez a su discurso. Naturalmente, reconocieron que el proceso de construcción del conocimiento era gradual y sufría considerables modificaciones desde la planeación hasta la consecución final del escrito. Los recursos materiales que empleaban para escribir eran generalmente diccionarios y traductores electrónicos, aunque la forma de emplearlos parecía estar dirigida más hacia un sistema eficiente de escritura que no siempre implicaba escribir todo en español durante el proceso de escritura, sino que también podía ser durante la edición del texto. Asimismo, se percibió una preferencia por la fluidez en el flujo del pensamiento plasmado en la escritura más que en la preferencia de escribir en una lengua en particular. Una vez que contaban con un escrito preliminar, los estudiantes reportaron acudir a diversas personas para

revisar sus textos, en general amistades o parejas, pero también manifestaron que esta ayuda se encontraba más bien focalizada a ciertos aspectos del texto.

Categoría 15-17. Descripción del proceso de escritura (incluye la fase de lectura, planeación, uso de recursos materiales y de la edición del texto con ayuda de terceros)

Investigador: ¿No se te dificulta escribir en español si lees en inglés?

E79: [...] Sí, el principio me costó mucho trabajo cambiar de uno al otro (refiriéndose a idiomas). Ya no tanto... [al principio] escribí todo en inglés así y luego la traducción... no, como ya no hago eso... escribo directamente en español... porque es como, es doble ¿no? Pero no es cuestión sólo de eso, es más, cuando fluye la escritura salen frases coloquiales, modos de expresión muy particulares ¿no?

Investigador: ¿Utilizas diccionarios u otros recursos para escribir?

E79: [...] Uso mucho mi traductor por teléfono, está... es muy x, pero bueno...

Investigador: Pero funciona.

E79: Sí, para palabras sí; para frases o... no, hace un desastre, pero para como palabras en... es que como una herramienta rápida, para obtener palabras importantes ¿no?

Investigador: ¿Le pides a alguien que revise tus textos?

E79: [...] Sí, este... pues entre mi novio y yo nos revisamos muchas cosas juntos ¿no? Este... [...] Y yo le ayudo con sus cosas en inglés y él en español.

E20: Primero leo las cosas que necesito. Eh... pongo... empiezo con una referencia que sé que es importante para el trabajo, algo así y luego lo que me pasa mucho es que pongo una idea, algo como más claro y en ese proceso se me ocurre algo y lo pongo más abajo. Entonces, cuando termino la idea, paso a la siguiente y puede ser un ir y venir entre lo que he leído y los puntos que necesito tocar. Después lo releo para ver si estoy repitiendo la información.

Investigador: ¿Cuándo escribes en la computadora, todo es en español?

E20: Sí, a menos que leí algo en inglés, entonces lo pongo en inglés y luego lo traduzco. Por ejemplo, ahora que estoy leyendo en alemán un libro, no sé si poner mis notas en alemán o español, pero hago lo que me viene más rápido. Los conceptos, como estoy hablando más en español, los hago en español. Si no la escribo en inglés, se me va la idea. Escribo un poquito más en inglés y luego la traduzco en español. Porque después siento mucho que se me va la idea. Si regreso a una nota, me cuesta trabajo recordar a qué se refería. Siempre es lo que me llega más rápido. Si lo conecto con una idea que ya tengo, aunque el artículo está en inglés, la escribo en español.

Investigador: ¿Utilizas diccionarios u otros recursos para escribir?

E20: Más que nada lo que hacía era buscar palabras o expresiones académicas en google. O, es que no me acuerdo bien, pero también en artículos iba identificando como empiezan, como concluyen, ciertas expresiones.

Investigador: ¿Le pides a alguien que revise tus textos?

E20: [...] En el escrito, poco a poco le iba dando mis trabajos a una amiga y ella me lo corregía. Más que nada me decía de la gramática. [...] Por ejemplo, mi tutor también me iba diciendo cosas. [...] sé que necesito que me chequen el trabajo, porque yo no lo noto [refiriéndose a cosas que escribe que no son claras para los lectores, pero que no sabe cómo autocorregir].

E15: [...] Cuando ya tengo más o menos las ideas, bueno, organizar el texto en partes, pero yo sé que cambia mucho según las personas, pero yo necesito escribir también para que mi pensamiento... ir más allá, más en el fondo. Entonces cuando tengo más o menos el esquema,

empiezo a escribir, pero voy remodelando un montón. Las frases, antes intentaba de una vez escribir una buena frase, ahora, como desde hace dos años, ahora escribo como vienen las ideas. Tengo un primer guion. [...] y ya luego voy formando, rescribiendo, como limpiando todo. [...] cuando hice ya todo eso de limpieza, pues lo mando leer a amigos hispanohablantes y también, bueno, mi compañero, mi novio es mexicano, entonces él ya lee todo, pobrecito (risas).

Investigador: ¿Utilizas diccionarios?

E15: [...] Eh... uso mucho *word reference*... eh... y ahí busco sinónimos. Porque además muchas veces escribo en francés porque es la primera cosa que me viene, pero no me cuesta tanto pasarlo al español. [...] Lo que más utilizo es el diccionario y también, sí seguro que leyendo textos en español, como tienes las ideas de una cierta forma, pues es más fácil.

Investigador: ¿Cómo es tu proceso de escritura, en español?

E84: [...] Sí, siempre.

Investigador: ¿A veces escribes las palabras en italiano y dices: bueno, después cuando revise...?

E84: [...] No, nunca, esto nunca. La frase siempre es en español, lo que se queda en mi computador siempre es en español. Nada más en mi cabeza pasa todo esto. [...] Eso es muy importante para mí, o sea, podría utilizar un vocabulario totalmente básico, pero es muy importante que la estructura sea original de la... [lengua de escritura].

Investigador: ¿Utilizas diccionarios para escribir?

E84: El diccionario, sí bueno, traductores porque son más rápidos, o sea me gustaría mucho tener tiempo para consultar un diccionario o... en papel. [...] muchas veces uso, el traductor de google ¿no? Es que es rápido, como que cuando escribo tengo flujos de pensamientos bastantes rápido y entonces, sí este... hago traducción.

E86: [...] Siempre empieza (quiso decir empiezo) en español, este... Yo creo que se me hace más difícil traducir las ideas si empiezo desde... Una amiga me dijo: "escribelo en inglés y mándamelo, ponlo en *google translate*, mándamelo y yo lo arreglo". Y le dije: "no pero... Todas las lecturas están en español, ¿cómo no voy a pensar? ni lo pienso en inglés [no se comprende claramente]. [...] Si estoy haciendo un trabajo escrito, como enseñé a mis alumnos para hacer sus exámenes de Cambridge, siempre empiezo con la planeación, con una planeación muy clara en español y de dónde voy a buscar ciertos fuentes de información y cómo lo voy hacer y lo planeo en, con... lápiz y con papel y luego empiezo a usar la compu y de repente, a veces, como tengo que intercambiar párrafos o hacer algo para que se enlacen mejor las ideas, o algo así, pero... pero siempre con este base en español.

Investigador: Si te llegan estas palabras en inglés, ¿qué haces? ¿Las buscas en el diccionario?

E86: [...] No, no, no, nunca. Odio, odio, odio interrumpir.

Investigador: Tu flujo de...

E86: [...] Exacto, exacto. Los meto en inglés y después, cuando estoy revisando mi trabajo, los busco en el diccionario. [...] la compu está más fácil con ciertas cosas como... Me corrige con la ortografía.

Investigador: ¿Le pides a alguien que revise tus textos?

E86: [...] Sí, este... como para esos trabajos finales, sobre todo como... Pido nada más que una persona me lo revise, pero... como no en cuestiones del contenido, sino del estilo.

E80: [...] me acostumbro a elaborar primero como un gráfico, un grafica con... los conceptos principales y luego desarrollar adaptarte y formarte una forma integral para elaborar luego un como ah... resumen, algo así, y luego voy a desarrollar cada parte concretamente.

Investigador: ¿Le pides a alguien que revise tus textos?

E80: [...] Ah, primero voy a corregir yo mismo, pero... como ya tengo amigos mexicanos si les dan tiempo voy a pedirle un favor ah... revisar un poco mi trabajo. Lo más frecuente voy a entregar a mi... a mi tutora para que me revisen y sería más profesional y mmm...y más familiar con tema que yo elaboro.

E4: [...] hago un plan. Que voy a tocar esos temas; hacer tantos capítulos, quiero tener capítulo que cubre este tema, este tema, este tema; este tema quiero hablar de eso; o usar tal autor, tal texto. Entonces, hago un plan y luego pues escribo.

Investigador: Tú estás escribiendo y te llegan palabras en francés. ¿Qué haces? ¿Lo escribes en francés y luego lo buscas en el diccionario?

E4: Pues sí tengo la idea. Si es un texto escrito, no siento que tengo que ir rápido porque... me va a tardar la idea ¿no? Voy escribiendo y a veces sí pongo la palabra en francés en paréntesis y luego voy a buscarlo en el diccionario después; a veces si estoy viendo en inglés, lo escribo en inglés y lo voy luego transformando o solamente retomando la idea. [...] O lo resumo en otra lengua y luego lo voy escribiendo en español. [...] Y hasta poner... en una frase... tener las tres lenguas. Es que son varios los procesos ¿eh? Depende mucho. Es un poco de todo, tengo... como muchas herramientas que voy usando de todas.

Investigador: ¿Utilizas diccionarios u otros recursos para escribir?

E4: Ah, diccionarios sí, todo el tiempo (risas) En el corrector... el diccionario español-español; pero el diccionario también... la traducción...

Investigador: ¿Le pides a alguien que revise tus textos?

E4: Sí, no es todo el tiempo, porque a veces... como por un trabajo ¿no? Algo así no, no he recurrido a los amigos. [...] En el primer semestre, segundo semestre, ya el tercer semestre estuve un poco más confiada.

E72: Bueno primero, bueno quizás antes que nada siempre hago como revisión de... de mi material.

Investigador: ¿De la literatura?

E72: ¡Aja! Cosas así. Luego empiezo hacer mis notas y luego me siento a escribir y... bueno, intento como escribir un par de horas sobre un capítulo, un tema lo que sea ¿no? Pero a veces sí es necesario recurrir al diccionario, si hay algo que no me suena bien o tengo una duda en cuanto a cómo se expresa esa idea o cómo se traduce algo, ¿no? Entonces sí, a veces tengo que consultar los diccionarios o incluso un traductor, si yo tengo algo ahí y me parece que algo anda mal quizás, voy a ponerlo en un traductor para ver la traducción que me sale; quizás en el google o algo así, para luego comparar con lo que tengo o la idea con... la idea que quiero expresar. Entonces, eso sí, a veces hago o sino, a veces pido a una amiga que me revise algo; que bueno eh... un hablante nativo va expresar eso así, lo va decir así y quizás me vayan diciendo: bueno no, hay que modificar esa idea acá. Cosas así hago.

Investigador: Sí pides ayuda.

E72: Sí, sí, sí y le digo: bueno, por favor me puedes echar un ojo aquí y dime si se entiende, si está bien dicho. Cosas así, pero generalmente para revisar todo un texto, no.

E74: Ante todo, leo y luego que he leído resumo todo para que se me fije la idea en mi mente; luego empiezo haciendo como un esquema y... empiezo a escribir no, siempre español, porque si ya empecé escribir en italiano pues se crea una confusión increíble. Entonces, pues escribiendo en español las primeras ideas; luego lo voy desarrollando y... normalmente corrijo mientras que hubiese escrito dos páginas, pues voy corrigiendo para que luego la corrección no sea tan dura al final del trabajo, si es muy largo. [...] algunas veces tengo unas dudas y para eso sirve *word translate* o *word reference*.

Investigador: ¿Le pides a alguien que revise tus textos?

E74: Claro, claro, también. Bueno, muchas veces no es que corrijan mi trabajo, si son muy largos, no quiero molestar. [...] tengo la suerte de tener dos compañeras que son españolas; entonces, muchas veces les pregunto a ellas, porque a veces traducen mejor que un traductor que no sabe de lo que está hablando. Entonces, creo es más útil, la verdad, preguntar a personas más que a traductores.

La retroalimentación en la lengua es importante para el aprendizaje continuo. En este sentido, la sensación general de los entrevistados fue que no era abundante por parte de los profesores, a menos que explícitamente ellos la solicitaran. Generalmente, quienes ofrecían más retroalimentación sobre la lengua eran los tutores y también los compañeros a quienes acudían para que les revisaran sus textos. De hecho, comentaron que a veces los profesores señalaban partes que no se entendían, pero de forma muy vaga explicaban qué es lo que no se comprendía.

Categoría 18. Retroalimentación que recibe sobre su escritura en español

E20: No, era solo un poco. Es que me decían que no se entendían algunas partes, pero otros no me corregían mucho.

Investigador: ¿Qué tan importante es la retroalimentación?

E20: Para mí es... los comentarios que me hacía mi tutor y cambiarlos a qué. Porque el problema es que me decían: “no se entiende”, pero yo no sabía qué. Entonces, no te daban opciones. Pero también es que en el trabajo final, el comentario era sobre el contenido. También cuando me leyeron la tesis final, una persona me dijo que unas cosas no eran claras, pero no me decían qué no era claro. También en la propuesta que hice para entrar al doctorado me dijeron que había cosas que no se entendían, pero no sabía qué. Por eso sé que necesito que me chequen el trabajo, porque yo no lo noto [refiriéndose a cosas que escribe que no son claras para los lectores]. [...] El problema es que como no me corregían mucho y nunca me decían nada, pues no sentía que estaba tan mal, que era suficiente. Por eso cuando me decían que algo no estaba bien, me quedaba siempre con la duda de dónde o en qué.

E15: Las ideas, pero la forma, no. [...] una maestra sí, pero porque creo que estada casada con un francés, entonces creo que es una francófila, le gusta mucho y lo hace porque sabe que sí necesito eso. Pero los demás no. [...]jella me dice: sería mejor así. Pero creo que es la única.

Investigador: ¿Ellos [los profesores] te ponen ahí mismo: oye, esto no se dice así en español, se dice así?

E84: Sí, sí, sí. Pero eso fue porque yo lo pedí a todos. O sea sí, sí, sí; al comienzo de la clase yo pedí que me... me corrigieran cuando, no sé, decía, sí, palabras poco correctas o que no se usaban mucho y entonces pedí como corregir.

E80: Ah... de todo tipo es que ya, al terminar mi tra... primer trabajo era un desastre y mi tutora a... me coligió [quiso decir corrigió] lo todo...

E4: No van a tomar el tiempo para darte este tipo de retroalimentación para nada, porque además no vuelvo a ver el examen. No me los entregaron. Y de todos modos, por lo general, hay exámenes que solo tengo la calificación final, no tengo ningún comentario ni en cuanto al contenido ni en cuanto a la manera de haberlo escrito. En trabajos largos, con algunos maestros que son clases más chiquitas, en una optativa que solo somos tres o cuatro, sí hay un poco de revisión gramatical, no estilística, sino gramatical.

E83: Ah... de los dos. Ah... obviamente de contenido; quieren ver que el contenido está muy bien; que están pasando más o menos, la mayoría; porque no son asignaturas de literatura o de escribir. La mayoría de las veces la documentación es sobre el contenido ah... pero también; sí, también te dan retroalimentación sobre el formato; si está bien. Ah... en el de estadística, estaba muy fijado el profesor en el formato que debías tener en la estructura y en la... hasta en la... en el estilo. Ah... porque el, la... la... el propósito era ah... replicar un reporte profesional, y en ese sentido todo esto era lo que iba a ayudarnos a aprender. Ah... y sí, enfocaba bastante también, no solamente en el contenido, pero también en el formato.

Investigador: ¿También de gramática te corrigen, este...?

E83: No, no, como buenos profesores de la UNAM, más que nada te dicen que no sales con errores, que revises mucho antes de entregarlo a mí. yo [refiriéndose al profesor] no quiero estar arreglando esta parte.

En relación con la reflexión sobre qué les ayudaría a mejorar su escritura académica, las opiniones vertidas se concentraron fundamentalmente en la experiencia al desarrollar trabajos escritos y en la retroalimentación personalizada de dichos trabajos, pero destacó el énfasis en la lectura como elemento que impulsa la escritura.

Categoría 19. Retroalimentación que recibe sobre su escritura en español

E79: [...] Yo digo como más práctica, sobre todo y... lo que me ayudó pues ¡huy! El camino ha sido largo (risas). Este trabajo me ayudó mucho. [...] Entonces, como después de un tiempo de estar leyendo, y también es agarrar un ritmo de escuela

E20: [...] Especialmente ahora, pienso que la retroalimentación que necesito es del trabajo que estoy haciendo. [...] Que me dijeran en un trabajo qué está mal y cómo se podría decir mejor, y eh... Que me dijeran cómo se puede decir mejor académicamente. Eh... Entonces, esa retroalimentación continua estaría padre.

E80: Mmm... creo que la lectura ah... al este nivel que yo cuento, ah... me ayuda más lectura de, por ejemplo, obras originales, novelas y desde allá puedo alimentarme ah... con el uso de palabras de una manera más mmm... nativa. Si logro tener en cuenta ah... los errores que yo suelo cometer y voy a dar a mmm de nuevo vistazo a libros de gramática, pero de versión original para ver si hay unos conceptos que mmm... a los que todavía no domino.

E68: Mmm... sí, escribir ensayos y que te... que te los corrijan, no que nada más se quede en la praxis, sino realmente, junto con un corrector de estilo, revisar el texto y decir: mira aquí puedes mejorar, ¿no? La... la corrección con alguien, nada más que te revise el texto y ya mira

aquí errores señalar, sino realmente hablar el texto, revisarlo juntos y con alguien que tenga un nivel admirable, superior que... que yo pueda aceptar las correcciones.

E82: Yo creo que leyendo se aprende mucho y... también como leyendo diferentes cosas, no sólo de... no solo la literatura científica sino también, literatura ¿no? Por ejemplo.

E74: [...] pues yo creo leer, seguir leyendo; porque veo que más leo, más puedo escribir y a veces, yo que sé, por ejemplo, eh palabras como... que... que funcionan como conexión mmm... No me salen naturales; pero más leo, más esta palabra me salen naturales, aunque la internalizo y luego salen ya... Entonces, sí, leer sí... a lo mejor... sí... como una guía un poco más específica por parte de los profesores cuando tienen que escribir. Eso sí creo que ayudaría porque a veces no... explican muy bien lo que [se] va hacer, y hay muchas, yo creo, discrecionalidad.

6.4.2.4. OTROS CONCEPTOS NO INCLUIDOS EN LAS CATEGORÍAS

La única categoría que resultó emergente tras el análisis de las entrevistas fue la 6: Toma clases de español o manifiesta una opinión al respecto. No obstante, hubo otros conceptos que se sistematizaron en la categoría 20 (otros conceptos) por ser eventos aislados. A continuación se destacan los comentarios más representativos de los estudiantes.

E73: [...] en Alemania, necesito, por este curso, necesito escribir un examen. Entonces es... diferente para mí, el maestro hace una calificación diferente sólo para mí.

La información mencionada por la estudiante 73 pone en evidencia que cuando los estudiantes se encuentran de intercambio académico, los requisitos académicos varían de acuerdo a lo que se les solicita en sus instituciones de origen. Por esta razón se decidió no incluir en el presente Estudio 3 a alumnos de licenciatura que se encontraban cursando un intercambio académico breve en la UNAM. Sin embargo, también había algunos alumnos de maestría que se encontraban haciendo intercambios académicos por un semestre o un año.

E08: [...] quiero seguir escribiendo en español. Justamente cuando regrese quiero meterme a un curso de preparación para el examen DELE y mejorar la parte gramática creo, la parte escritura también y sí, pues quiero, aunque no... a pesar que no tenga la oportunidad creo como que ah... oportunidad de escribir en español como en la universidad. Entonces, creo que va a ser muy útil para los textos en este curso y luego... luego, no sé. En cualquier trabajo que debo entregar, porque yo quiero como seguir avanzando mi español, aparte mi inglés y bueno, pues... Yo quiero ir avanzando, pero no sé. La parte de escritura, creo que va a ser como la más olvidada.

Investigador: ¿Alguna vez te sentiste desmotivado porque sabías que no ibas a escribir más en español?

No. No. Si hubiera sido desmotivado, no habría llegado a México. Es que yo lo veo como la experiencia muy buena. La oportunidad, porque yo quiero enfocarme en Relaciones Internacionales en América Latina. Yo voy a necesitar mi español de cualquier modo. Entonces, nunca he tenido este... impresión que era trabajo... (piensa).

Investigador: En vano.

E8: ¡Exacto! En vano.

El estudiante 8 expresó un comentario de suma relevancia para el aprendizaje de un idioma: la motivación. Aunque él se encontraba de intercambio por tan solo un semestre, y quizás veía a corto plazo un escaso uso de su español, su visión del uso del español a largo plazo era clara. Valdría la pena incluir una categoría en futuras investigaciones que se relacionara con la motivación que tienen para seguir mejorando su escritura académica, así como si piensan seguir usando la escritura académica en español en el futuro.

E79: [...] Ya tengo mucha confianza de que como... bueno, ya acepto mucho el que soy extranjera, nunca voy a ser mexicana (risas), nunca me van a aceptar 100% y eso... es parte del contrato ¿no? (risas).

Esta visión de la estudiante 79 resulta también interesante para seguir explorando, pues había quienes mencionaban que en su proceso de aculturación se olvidaban de ser extranjeros y se enfocaban en la lengua y la cultura mexicana; no obstante, el comentario anterior de la alumna pone en claro que ella se percibe y asume como extranjera, sin que esto le genere un conflicto. Sería bueno indagar sobre qué tan aculturados se encuentran los estudiantes y cómo sienten ellos que esto afecta su desempeño académico; en este caso, de escritura.

6.5. DISCUSIONES, LIMITACIONES Y FUTURAS INVESTIGACIONES

En este apartado se abordan las discusiones de los dos estudios que se condujeron para dar respaldo a la inferencia de la extrapolación. La sección de las discusiones se encuentra dividida en dos partes, de acuerdo con las dos aseveraciones que se establecieron para la inferencia de la extrapolación: la correlación de la PE del EXELEAA con otros instrumentos de evaluación de escritura académica, y las habilidades de escritura académica de los estudiantes de posgrado no hispanohablantes. Más adelante se ofrecen las limitaciones del estudio de extrapolación y también se presentan alternativas y sugerencias para futuras investigaciones.

6.5.1. PRIMERA PARTE: LA CORRELACIÓN DE LA PE DEL EXELEAA CON OTROS INSTRUMENTOS DE EVALUACIÓN DE ESCRITURA ACADÉMICA

El objetivo de la primera parte del estudio consistió en correlacionar los resultados que obtuvieron los estudiantes en la PE del EXLEAA con una encuesta de autoevaluación y también con una entrevista semiestructurada. Uno de los objetivos fue buscar fuentes alternativas de comparación debido a los problemas que se han suscitado al tratar de comparar los resultados de un ECLAA con las calificaciones que obtienen los estudiantes en sus estudios universitarios, o con los textos que redactan en el ámbito académico. En este sentido, se buscaron medios alternos de escritura académica que pudieran ser comparables con los resultados de la PE.

La primera aseercción que se planteó fue que los resultados obtenidos por los estudiantes en la PE del EXELEAA correlacionan de forma positiva y significativa con otros instrumentos de evaluación de escritura académica en español. Como se pudo ver en la primera parte de este estudio, la PE escrita correlacionó moderadamente con la encuesta de autoevaluación de las habilidades de escritura académica y correlacionó fuertemente con los resultados de la evaluación de la entrevista. Los hallazgos de la encuesta van de la mano con los resultados de correlaciones moderadas que Powers (2015); Powers et al. (2010), Sinharay et al. (2009) y Weigle (2010) han reportado. Sin embargo, una de las ventajas del presente estudio consistió en que el cuestionario se circunscribía exclusivamente a la habilidad de escritura, a diferencia de los autores antes mencionados, quienes han correlacionado las distintas habilidades de lengua (i.e., comprensión de lectura, auditiva, expresión oral y escrita) con cuestionarios y entrevistas. Otra de las ventajas del presente estudio consistió en el aspecto experiencial; es decir, los estudiantes encuestados se encontraban matriculados en algún posgrado de la UNAM y, por lo menos, habían concluido su primer semestre académico. Como bien menciona Suzuki (2015), cuando se les solicita a los estudiantes que autoevalúen lo que son capaces de hacer, los resultados son mucho más certeros cuando se encuentran desempeñando dichas tareas; de lo contrario, sus valoraciones resultan en meras especulaciones.

No obstante, aunque se trató de controlar las variables antes mencionadas, la correlación entre la encuesta y la PE fue moderada. Esto deja abiertas algunas interrogantes de por qué la correlación no fue mayor. Algunas posibles explicaciones se encuentran en la experiencia, el nivel y tipo de estudios, la lengua materna, los enunciados contenidos en el descriptor de escritura académica y probablemente en características individuales de los estudiantes. En

cuanto a la experiencia, el tiempo de estancia en México y de escribir en un contexto académico en español osciló entre tres meses y ocho años. En este sentido, los estudiantes con menor experiencia pudieron haber subestimado o sobreestimado sus habilidades de escritura académica, mientras que los estudiantes con mucha más experiencia, podrían haber sido más objetivos, aunque también modestos en sus evaluaciones. Esta posible interpretación va de la mano con resultados reportados por Suzuki (2015) en estudiantes chinos que se encontraban estudiando en Japón. El autor encontró que quienes contaban con mayor experiencia académica eran más objetivos, aunque modestos, en sus autoevaluaciones sobre su nivel de dominio de la lengua meta: el japonés.

Otra posible razón de discrepancia entre los resultados de la encuesta y de la PE del EXELEAA se encuentra en el tipo de estudios y en el nivel (maestría o doctorado), puesto que los estudiantes de doctorado, al no contar con materias curriculares, enfocan la mayor cantidad de su trabajo académico a la lectura y a la redacción de su tesis. El nivel de exigencia de escritura de estos estudiantes es el de una tesis doctoral. Por lo tanto, este grupo podría percibir que el dominio de sus habilidades de escritura era más bajo que el de aquellos que desarrollan otro tipo de trabajos académicos.

Adicionalmente, la encuesta incluía una amplia gama de habilidades académicas, de las cuales unas son más representativas de ciertas disciplinas académicas que otras. Más aún, como se pudo observar en el Estudio 2 de esta presente investigación, hay numerosas habilidades de escritura académica que no son evaluadas por la rúbrica analítica, debido al tipo de tareas discretas que se incluyeron en la PE del EXELEAA. Estas habilidades se relacionan principalmente con la construcción del conocimiento y la incorporación del conocimiento previo en la redacción de trabajos. En la encuesta, sin embargo, sí se les pedía a los estudiantes que identificaran el nivel de dominio que consideraban tener en cada una de estas habilidades.

Finalmente, también las características personales del estudiante influyen. Como bien mencionan Sinharay et al. (2009), los estudiantes de ciertas culturas tienden a sobreestimar sus habilidades, mientras que otras culturas, a subestimarlas. Grosso modo, todas estas variables pudieron haber incidido negativamente para que la correlación entre la encuesta y la PE no fuera mayor.

En cuanto a la entrevista semiestructurada, la correlación con la PE fue fuerte. No se sabe aún de estudios en evaluación de lenguas que hayan abordado la extrapolación desde esta perspectiva. La conversión de datos cualitativos a cuantitativos no fue sencilla; de hecho, este tipo de estudios es aún incipiente en las ciencias sociales (ver Srnka y Koeszegi, 2007). La construcción de la rúbrica analítica partió de las categorías que se evidenciaron en las entrevistas; sin embargo, no se pudo evaluar a todos los estudiantes en todas las categorías, puesto que no siempre aportaron información relacionada con cada una de ellas.

Adicionalmente, existen algunas fuentes de sesgos al momento de evaluar las entrevistas. Por ejemplo, los estudiantes orientales generalmente presentaban dificultades durante la entrevista para expresar sus ideas o para ahondar en las preguntas que se les hacían, aun cuando su nivel de escritura fuera adecuado. Esto podría generar una percepción errónea en el evaluador. En cambio, los estudiantes con lenguas afines al español por lo general interactuaban y detallaban más sus procesos de escritura. Si bien las entrevistas se encontraban ya transcritas al momento de evaluarlas mediante la rúbrica analítica, se dejó abierta la posibilidad de escuchar los audios para evaluarlas. Por esta razón, resultó de suma importancia que las entrevistas fueran evaluadas por al menos dos personas.

Por último, cabe recordar que las tareas discretas incluidas en la PE del EXELEAA no miden las habilidades relacionadas con la construcción del conocimiento ni tampoco aquellas relacionadas con el conocimiento previo; no obstante, los resultados de correlación con la entrevista son altos. Esto da indicios de que, pese a la falta de incorporar tareas integradas en el EXELEAA, las tareas incluidas en la PE son de utilidad para evaluar las habilidades de escritura de los prospectos universitarios. Sin embargo, como se pudo observar, los estudiantes clasificados en el grupo intermedio bajo en la entrevista habían sido clasificados en el grupo intermedio alto en la PE del EXELEAA. Esto significa que las tareas de escritura del EXELEAA resultaron relativamente más sencillas que las habilidades de escritura que reportaron tener los estudiantes durante la entrevista. Valdría la pena, entonces, incorporar tareas más retadoras, como las integradas, en aras de explorar si éstas establecen un mejor equilibrio entre lo que se solicita en el examen y lo que los estudiantes demuestran en su quehacer académico.

6.5.2. SEGUNDA PARTE: LAS HABILIDADES DE ESCRITURA ACADÉMICA DE LOS ESTUDIANTES NO HISPANOHABLANTES

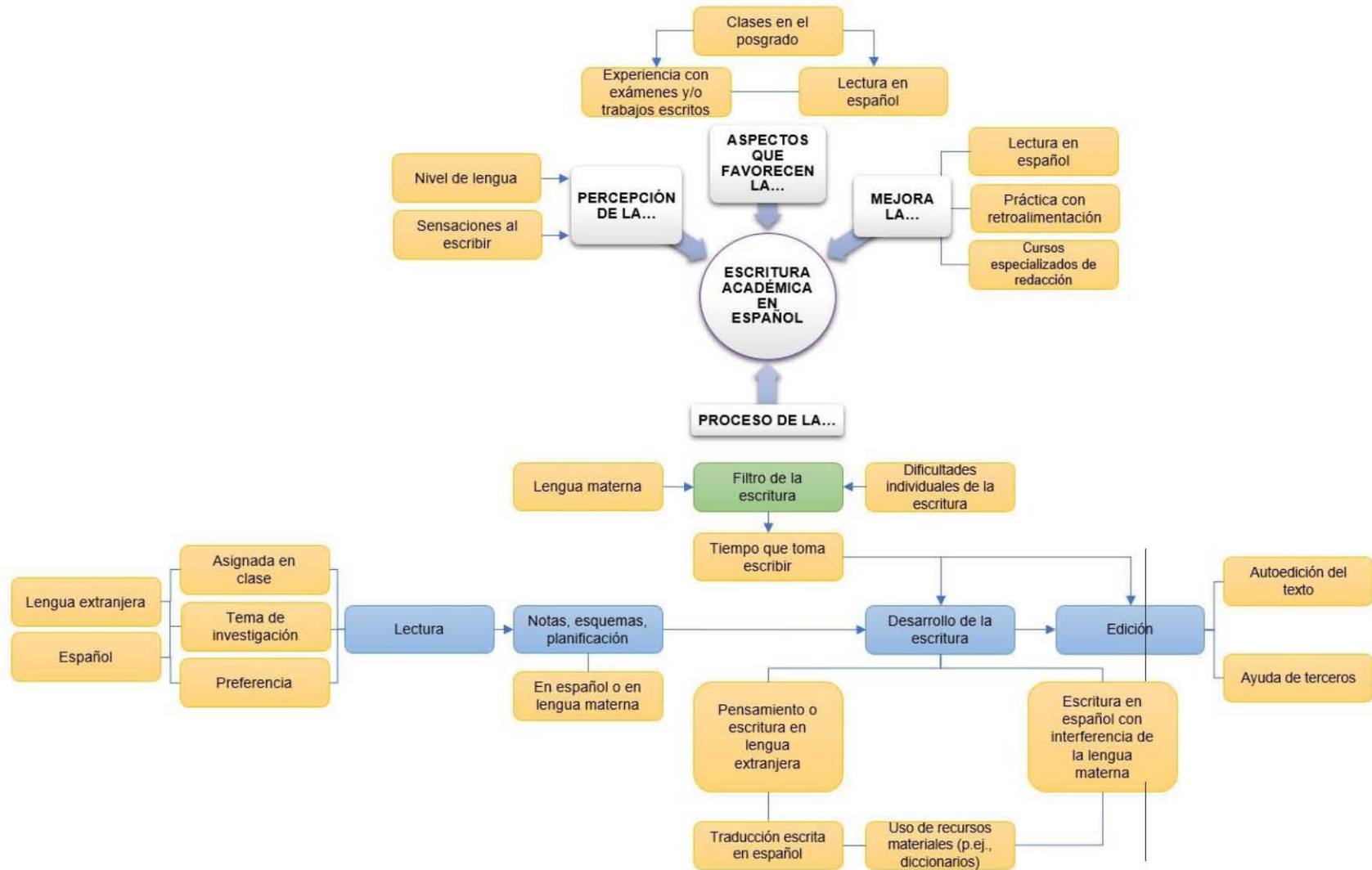
En la segunda asección de este estudio se estableció que los estudiantes más proficientes demostrarían un nivel de escritura académica superior en comparación con aquellos menos proficientes. Los estudiantes fueron clasificados en alguno de los tres siguientes grupos: intermedio bajo, intermedio alto y avanzado –de acuerdo con el resultado que obtuvieron en la entrevista–. A continuación, se presentará el modelo de escritura académica que se diseñó a partir del análisis de las entrevistas.

6.5.2.1. EL MODELO DE ESCRITURA ACADÉMICA EN ESPAÑOL COMO SEGUNDA LENGUA

Uno de los principales objetivos de la investigación cualitativa es el desarrollo de teorías o modelos que permitan explicar de qué forma se relacionan los datos observados (Bazeley, 2009). Hasta el momento no existe teoría alguna que explique el proceso de escritura (Knoch, 2011; Weigle, 2002), y los modelos existentes generalmente se enfocan al proceso de escritura en lengua materna desde el punto de vista cognitivo (Bereiter y Scardamalia, 1987; Hayes, 1996; Hayes y Flower, 1980; Kellog, 1996). Weigle (2002) menciona que a diferencia de la escritura en lengua materna, en segundas lenguas, existen dos componentes independientes aunque interrelacionados entre sí: el dominio de la segunda lengua y el dominio de las habilidades de escritura. Según la autora, la dificultad al escribir se presenta cuando el estudiante se concentra más en la lengua que en el contenido. Esto se pudo ver claramente en el análisis de las entrevistas. De hecho, los estudiantes del grupo intermedio bajo se encuentran aún por debajo del umbral mínimo para escribir eficazmente en español. De ahí que, sus sensaciones al escribir sean particularmente negativas. Por último, otra de las limitaciones de los modelos de escritura radica en que se restringen exclusivamente al proceso de escritura y no permiten comprender de qué forma el entorno –en este caso el académico– se relaciona con las habilidades de escritura.

La Figura 6.9 presenta el modelo de escritura académica en español que se derivó del análisis de las entrevistas. El modelo se encuentra compuesto por un componente principal: el proceso de escritura académica, y por tres componentes más asociados a él: la autopercepción de la escritura, aspectos que favorecen la escritura y reflexiones sobre qué podría mejorar la escritura.

Figura 6.9: Modelo de escritura académica en español como segunda lengua (Desarrollado por el investigador)



El punto central del modelo se encuentra determinado por el proceso de escritura. Los estudiantes, con base en su experiencia al escribir en el contexto universitario, tienen ciertas percepciones de su escritura académica en español. El proceso de escritura se ve también favorecido por determinados factores que influyen en la escritura (p.ej., si toman clases en el posgrado o de lengua, su experiencia con trabajos y/o exámenes académicos y con la lectura en español). Por último, la reflexión sobre qué puede mejorar la escritura constituye también un componente esencial asociado al desarrollo de la escritura académica.

A continuación, se hará una descripción detallada de cada uno de los cuatro componentes del modelo de escritura académica. En primer lugar, se abordará el proceso de escritura. Como se puede observar en la Figura 6.9, este proceso –determinado por el nivel de dominio de la lengua meta y la influencia de la lengua materna– comienza generalmente con la lectura, la toma de notas y la elaboración de mapas mentales y/o esquemas, continúa con el desarrollo propio de la escritura con ayuda de recursos materiales y concluye con la edición del texto.

6.5.2.1.1. EL PROCESO DE ESCRITURA EN ESPAÑOL COMO LENGUA EXTRANJERA

El proceso de escritura en español de los estudiantes no hispanohablantes se encuentra tamizado por dos filtros importantes: la influencia de la lengua materna y las dificultades individuales con la lengua meta. La lengua materna parece tener un papel importante en el proceso de escritura. Naturalmente, lenguas lejanas al español (p.ej., polaco, checo, chino, japonés, coreano) representan sistemas de escritura distintos que en muchas ocasiones no tienen relación alguna con la estructura del español. Esto fuerza a los estudiantes de estas lenguas a aprender un sistema de escritura completamente diferente; en consecuencia, su lengua materna no les permite acceder fácilmente al sistema de escritura del español. Por el contrario, las lenguas maternas afines al español (p.ej., italiano, francés e inglés) sí facilitan el proceso de escritura, pues la estructura de la lengua es similar. No obstante, los estudiantes con lenguas afines al español mencionaron que esto no siempre era favorable, pues aunque la estructura de las lenguas era similar, no era igual.

El otro filtro dentro del proceso de escritura se encuentra determinado por el nivel de lengua de cada uno de los estudiantes. Pese a que todos los estudiantes de los tres grupos afirmaron tener deficiencias lingüísticas al momento de escribir en español, estas son de diversa naturaleza. Por ejemplo, a nivel microestructura, los tres grupos mencionaron que les

hacia falta vocabulario especializado, académico y formal; y presentaban problemas de gramática, sintaxis y puntuación. Los problemas a nivel oración se vieron mucho más acentuados en el nivel intermedio bajo y menos en el grupo intermedio alto y avanzado. En estos últimos dos grupos, la mayor dificultad está representada por la falta de vocabulario especializado, académico y con las convenciones formales que se requieren. En el nivel de macroestructura, no reportaron problemas sustanciales, salvo la dificultad de saber qué se espera cuando se les pide un trabajo de un género discursivo en particular; por ejemplo, un ensayo, una reseña o un resumen crítico. Es decir, la manera de abordar distintos géneros no siempre es igual entre lenguas distintas.

Una vez que la capacidad de escritura se ve filtrada por la lengua materna y el nivel de dominio de la lengua meta, ésta se ve reflejada en el proceso de escritura. El proceso de escritura no parte propiamente de la escritura, sino desde la misma lectura. Es en este momento cuando el estudiante comienza a generar ideas y conceptos sobre los cuales escribirá posteriormente. También en este momento se apropia del léxico, de las estructuras lingüísticas y convenciones académicas del español, claro, cuando la lectura está ese idioma.

En el caso de los estudiantes de maestría, los docentes asignan lecturas obligatorias en una lengua determinada. Dependiendo de la disciplina académica, esta lectura puede ser en español, en lengua extranjera o en ambas. Los estudiantes de doctorado, por el tipo de investigación, seleccionan lecturas en lengua extranjera o en español o en ambas. Estos dos tipos de variables son externas; es decir, el estudiante no interviene en la selección del idioma de las lecturas, sino que se encuentran determinadas por el docente o por el tipo de investigación. Sin embargo, los estudiantes también eligen material de lectura adicional para enriquecer su formación académica. En este caso, la selección del idioma se encuentra determinada por motivos personales o incluso estratégicos. Por ejemplo, hay quienes prefieren leer en inglés porque les parece que la lectura es más clara, la información se encuentra más actualizada y muchas veces hay más material disponible en la Internet. Por el contrario, ciertos estudiantes optan por la lectura en español porque les ayudará eventualmente al proceso de escritura, o simplemente la eligen en un idioma por gusto. También hay estudiantes a quienes les resulta indiferente uno u otro idioma. Finalmente, hay quienes optan por un idioma particular porque les resulta más sencillo.

Una vez que se lee el material, los estudiantes toman notas o elaboran fichas informativas que les servirán posteriormente para la escritura. También durante este proceso se elaboran esquemas y organigramas de lo que desean escribir e incluir en cada uno de los apartados. Las notas también provienen de sus apuntes de clase o conferencias a las cuales asisten. El idioma de estas notas varía considerablemente entre los grupos. Los estudiantes del grupo intermedio bajo tienden a escribir mayoritariamente en otra lengua; de hecho, resaltaron que cuando no sabían cómo escribir algo, lo escribían en su lengua materna para no confundirse y recordar el concepto. Una estudiante incluso mencionó que escribía fonéticamente en coreano. Parece ser que el proceso de adquisición de la lengua extranjera transita por este primer estadio de escritura en lengua materna y, conforme se incrementa el dominio, se escribe más en español. Un estudiante del grupo intermedio alto destacó que al principio, cuando su nivel de español no era suficiente, escuchaba en español y escribía en francés, aunque recalcó que después de tres años de doctorado, no necesitaría hacerlo más de ese modo.

Los estudiantes del grupo intermedio alto y avanzado también mencionaron que sus notas se encontraban mezcladas en español y lengua materna, aunque matizaron que esto se debía al idioma de las lecturas o al idioma en que se hablaba. En el grupo avanzado parece ser que la razón por la cual utilizan la lengua extranjera se debe más bien a ciertas estrategias de aprendizaje y de escritura. En todo caso, el uso de la lengua materna resulta ser una valiosa herramienta para recordar conceptos importantes que requieren ser bien afianzados.

Una vez que los estudiantes cuentan con sus notas, esquemas y diagramas de planificación de la escritura, comienza el desarrollo de la escritura en computadora. Dependiendo del nivel de lengua del estudiante o de sus preferencias de escritura, el uso de recursos materiales (p.ej., diccionarios, traductores, foros de discusión) es simultáneo a la escritura o posterior durante la fase de edición. Los filtros de la lengua materna y de las dificultades al escribir en español repercuten directamente en el tiempo que toma este desarrollo de la escritura. En cuanto mayores son las dificultades lingüísticas y más distante es la lengua materna, mayor será el tiempo que le tome escribir al estudiante (incluso más del doble de lo que le tomaría hacerlo en lengua materna). Por el contrario, este tiempo disminuye conforme incrementa el nivel de dominio lingüístico del estudiante y la lengua materna se acerca más al español. De hecho, el tiempo de escritura de los estudiantes del grupo avanzado

resultó ser prácticamente similar al tiempo que toma la redacción en lengua materna, aunque ligeramente mayor durante el proceso de edición.

En relación con el desarrollo de la escritura, se percibieron diferencias sustanciales entre los tres grupos. Los estudiantes con menor dominio reportaron un proceso de escritura prácticamente de traducción literal al español; ya sea de manera simultánea (i.e., se piensa primero en otra lengua y se intenta traducir al español), o bien, escriben primero el texto en lengua extranjera y después traducen el texto al español. Cuando los estudiantes del grupo intermedio bajo hacen el esfuerzo por escribir directamente en español, la falta de recursos lingüísticos bloquea el flujo de pensamiento, pues los recursos cognitivos se encuentran centrados en la lengua y no en el contenido (Weigle 2002). Por esta razón, estos estudiantes prefieren escribir en otra lengua y después hacer la traducción. De esta forma, el pensamiento no se ve interrumpido por las deficiencias en la lengua meta. Adicionalmente, estos estudiantes son dependientes en todo momento de los recursos materiales para escribir en español (diccionario, foros de discusión o traductores electrónicos), pues sin ellos, su proceso de escritura se ve imposibilitado. Otra de las dificultades con el uso del diccionario es que no logran comprender el uso de las distintas acepciones de una palabra.

Los integrantes del grupo intermedio alto cuentan con un proceso de escritura en español más desarrollado. Escriben sus textos en español, aunque algunos intercalan palabras o frases en otro idioma. En general, su proceso de escritura no es tan fluido porque su pensamiento, si bien es mayoritariamente en español, presenta algunas lagunas lingüísticas (i.e., gramaticales, sintácticas, léxicas) que les impiden expresar sus ideas por escrito de forma continua y sin interrupciones. Al no encontrar la palabra, frase o estructura gramatical, la escriben en otro idioma, pero el flujo de pensamiento se ve parcialmente interrumpido. También hay quienes desean escribir todo su texto de una vez en español, para lo cual invierten tiempo de búsqueda en traductores electrónicos, que va en detrimento del flujo de ideas. Aparentemente este grupo no es tan dependiente del diccionario/traductor electrónico, pero lo ocupan con relativa frecuencia para tener más riqueza léxica.

Finalmente, el grupo avanzado denota un desarrollo de escritura bastante fluido. Su pensamiento en español se encuentra mucho más consolidado. De vez en cuando escriben alguna palabra en otra lengua o emplean traductores electrónicos para buscar sinónimos o léxico que resulte más académico o especializado. En general, cualesquiera de las estrategias

que emplean va en favor del flujo del pensamiento. Es decir, con tal de no perder una idea, pueden mezclar información en español o en otra lengua.

El proceso de edición de los textos es variado. Hay quienes lo hacen paulatinamente y otros más cuando ya cuentan con un borrador preliminar de sus escritos. El proceso inicial de edición es llevado a cabo por los mismos estudiantes. Posteriormente, es común que acudan a terceros para que les revisen sus textos. En el caso de los estudiantes del grupo intermedio bajo, se pudo apreciar que no contaban con estrategias eficientes de ayuda en la edición de sus textos. O no le piden a nadie que les corrijan sus textos, o son completamente dependientes de la revisión otros. Los estudiantes del grupo intermedio alto y avanzado parecen recurrir de igual forma a la ayuda de terceros para revisar sus textos. En estos casos, los estudiantes son más selectivos y solicitan la revisión de ciertas partes del texto. Esto demuestra también que están más conscientes de dónde tienen problemas para expresarse. Muchos de los estudiantes entrevistados tienen parejas sentimentales hispanohablantes. En estos casos, la pareja resulta un valioso medio de corrección de trabajos escritos, lo cual ayuda también a la mejora paulatina de la escritura académica. Otra figura primordial es la del tutor, pues éste juega un papel importante en el proceso de corrección de la lengua, dado que, en el contexto académico, prácticamente es la única figura académica que hace correcciones (en términos del uso de la lengua) del trabajo de investigación de los estudiantes de posgrado. Naturalmente, cuando el tutor es nativohablante del español, participa mucho más en el proceso de corrección del estudiante que cuando se trata de tutores no nativohablantes.

6.5.2.1.2. LA PERCEPCIÓN DE LA ESCRITURA ACADÉMICA

Con respecto a su autopercepción del nivel de escritura en español y las emociones que experimentan al escribir, se pudo apreciar que, en general, en el grupo intermedio bajo estas sensaciones de escritura son negativas y la percepción del nivel de escritura es bajo. El principal problema es que las sensaciones de minusvalía sobre la escritura académica podrían afectar considerablemente la motivación y el desempeño académico del estudiante, poniéndolo en riesgo de abandonar sus estudios. Sin embargo, conforme incrementa el nivel de dominio en la lengua meta, estas percepciones se modifican. En el grupo intermedio alto, los comentarios se enfocaron en la dificultad que representa escribir en español y algunos reportaron una mejoría en comparación a cuando iniciaron sus estudios. De los resultados del grupo avanzado, destaca la variedad de respuestas sobre su nivel de escritura. Hubo quienes mencionaron que éste era básico o intermedio. Lo interesante fue que quienes mencionaron

que su nivel de escritura era avanzado, recalcaron que aún podía mejorar. De igual forma, con respecto a las sensaciones, si bien en general fueron de gusto y confianza en el grupo avanzado, también matizaron que no se comparaba a la escritura en lengua materna. Esto evidencia que la habilidad de escritura en lengua extranjera siempre constituye un reto para los estudiantes, incluso en niveles avanzados.

6.5.2.1.3. ASPECTOS QUE FAVORECEN LA ESCRITURA ACADÉMICA EN ESPAÑOL

Sobre los cursos de español, sólo en el grupo intermedio bajo se destacó la importancia de tomar clases de español antes de tomar clases en el posgrado. En general, los estudiantes comentaron que los cursos de gramática avanzada o de redacción enfocados a la revisión de textos académicos serían deseables. En relación con las clases que toman en el posgrado, parece ser que esto más bien depende del nivel de estudios y del posgrado. Es decir, los estudiantes de maestría se ven obligados a tomar clases curriculares, lo cual les ayuda significativamente a mejorar paulatinamente su escritura académica. En cambio, algunos de los estudiantes de doctorado no se ven obligados a tomar clases en sus posgrados, lo cual imposibilita que poco a poco mejoren su producción escrita. No obstante, algunos estudiantes del grupo intermedio alto y avanzado expresaron tomar materias adicionales como estrategia para mejorar su escritura, su nivel de español en general, profundizar en sus temas de investigación o simplemente para entablar relaciones sociales.

De manera generalizada, los estudiantes mencionaron que casi no habían tenido exámenes escritos en sus posgrados. Esto va de la mano con las prácticas de evaluación docente (ver Capítulo 4, Estudio1), pues de los estudiantes entrevistados, solamente tres pertenecían al doctorado en ciencias de la tierra; es decir, la gran mayoría de los entrevistados eran alumnos del área de Ciencias Sociales, Humanidades y Artes. En estas áreas del saber, los exámenes en clase no representan métodos tradicionales de evaluación. Sin embargo, si eventualmente se replicara el estudio con una muestra más amplia y representativa de las áreas Físico-Matemáticas e Ingenierías, y de Ciencias Biológicas y de la Salud, quizás esta categoría mostraría diferencias entre los grupos menos proficientes y con mayor dominio lingüístico. Contrariamente, la gran mayoría de los tres grupos mencionó que tenían experiencia con trabajos escritos (p.ej., trabajos y tareas cortos, trabajos de fin de curso, ensayos, capítulos de tesis), excepto los estudiantes que apenas se encontraban por concluir su primer semestre académico. Estos estudiantes, como se especuló desde el inicio, no podían describir con detalle su proceso de escritura, pues no tenían casi experiencia. Se especulaba

que los estudiantes que tuvieran mayor experiencia con trabajos escritos se encontrarían en los grupos intermedio alto y avanzado. Sin embargo, hubo estudiantes que se encontraban redactando sus tesis doctorales en español –desde hacía un año, por lo menos– y que pertenecían al grupo bajo.

6.5.2.1.4. LA REFLEXIÓN SOBRE LA MEJORA DE LA ESCRITURA ACADÉMICA

Finalmente, sobre qué podría mejorar la escritura académica, la gran mayoría de los tres grupos reconoció que ésta se favorece a través de la lectura y también a partir de la práctica escrita siempre y cuando haya corrección individual de textos. Los estudiantes mencionaron que a través de la lectura en español se adquiere vocabulario, expresiones y frases académicas, además de que también se aprenden estructuras lingüísticas y propias del ámbito académico. En cuanto a la corrección, esta podría ser por docentes o tutores del posgrado, compañeros de clase o en clases de redacción enfocadas específicamente a la revisión de textos académicos. En suma, si bien esta categoría aportó información muy similar entre los tres grupos, es particularmente útil.

6.5.3. LIMITACIONES Y FUTURAS INVESTIGACIONES

El presente estudio presenta algunas limitaciones. En primer lugar, los participantes del Estudio 3 son estudiantes de maestría y doctorado. Inicialmente se había contemplado la participación exclusiva de estudiantes de maestría; sin embargo, debido a que la muestra se reducía considerablemente, se decidió incluir también a estudiantes de doctorado. Los estudiantes de doctorado presentan un perfil de escritura académica diferente a los de maestría. Los estudiantes de maestría cursan materias curriculares durante sus dos años de estudio, por ende, su experiencia con trabajos académicos es mucho más diversa que la de los estudiantes de doctorado, quienes mayoritariamente se enfocan a la redacción de su tesis. En este sentido, el proceso de escritura de una tesis doctoral es mucho más riguroso que los trabajos académicos que presentan los estudiantes de maestría.

Otra de las limitaciones se relaciona con la procedencia académica de los participantes. Aunque en la entrevista hubo tres participantes de Ciencias de la Tierra, en general todos los estudiantes del Estudio 3 provenían de Ciencias Sociales, Humanidades y Artes. Debido a que la participación al estudio fue voluntaria, esta variable no se pudo controlar. Ya que la gran mayoría de los estudiantes que participaron en el estudio contaban con un certificado de dominio expedido por el CEPE, fue difícil incentivar su participación en el estudio.

Adicionalmente, pese a que el estudio con los estudiantes se llevó a cabo al final del semestre, con el fin de que por lo menos contaran con un semestre de experiencia académica, algunos de quienes se encontraban cursando apenas el primer semestre no habían entregado aún sus trabajos académicos finales. Sin lugar a dudas, esto limitó el grado de experiencia con trabajos académicos tanto para contestar la encuesta como dentro su participación en la entrevista.

Finalmente, la muestra que participó en el Estudio 3 pudo haber sido mayor. Debido a que en la primera sesión las entrevistas se condujeron junto con el taller de escritura académica una semana después del examen, el número de participantes fue reducido. Por tal razón, con el fin de captar a un número mayor de participantes en las entrevistas, en la segunda sesión se decidió entrevistar a los estudiantes el mismo día del examen. Esto permitió entrevistar a un número mayor de estudiantes, pero representó una limitación en cuanto al tiempo que se le destinó a cada uno de ellos. Consecuentemente, la profundización en el tema no fue tan rica como en la primera sesión. En el caso de la encuesta, habría ayudado que contestaran obligatoriamente la encuesta antes de resolver el examen. Eso habría garantizado una participación mayor. Debido al tamaño de los participantes en la entrevista, no se pudo investigar la calidad de la rúbrica mediante el análisis estadístico MFRM. Linacre (citado en Barkaoui, 2014) sugiere que por lo menos se cuente con 30 examinados. El análisis permitiría identificar qué categorías funcionan adecuadamente y también ayudaría a determinar la consistencia de los evaluadores.

Estudios posteriores podrían abordar también la participación de los tutores. De manera puntual, se podría entrevistar a los asesores de los estudiantes no hispanohablantes respecto del dominio de escritura académica de sus tutorandos. De forma alternativa, también se podría diseñar y aplicar una encuesta a los tutores. Sin embargo, para poder llevar a cabo un estudio de este tipo, se debería garantizar que el tutor ha revisado el progreso de la tesis del estudiante durante cierto tiempo; es decir, sería importante que conociera bien su dominio de escritura. En este tipo de estudios se debería excluir a los docentes (cfr. Weigle, 2010). El problema de incorporar a los docentes radica en que es difícil que puedan conocer con detalle el dominio de escritura de un estudiante –puesto que no trabajan de cerca con cada uno de ellos– adicionalmente, no se podría saber hasta qué punto el trabajo del estudiante ha sido editado con ayuda de terceros.

6.6. CONCLUSIONES

El fin de la evaluación académica no está en el puntaje que se obtiene de los exámenes, sino en que éstos sirvan como un instrumento que permita tomar decisiones justas con respecto al desempeño de los candidatos que desean ingresar al ciclo universitario. Por esta razón, pese a que los estudios de extrapolación de los resultados de un ECLAA conlleven un intrincado proceso que hasta la fecha no había logrado esclarecer de qué forma se evidencia el resultado obtenido en un examen con el mundo real, éstos son necesarios para justificar la naturaleza misma de la evaluación.

En este estudio se abordó la inferencia de extrapolación de los resultados de la PE del EXELEAA al “mundo real”. Para respaldar esta inferencia, se enunciaron dos aseercciones. En la primera de ellas, relacionada con la correlación de la PE escrita del EXELEAA con otros instrumentos que evalúan la escritura académica de los estudiantes, se evidenció que la PE correlacionó moderadamente con la encuesta de autoevaluación de habilidades de escritura académica y correlacionó fuertemente con los resultados de la evaluación de la entrevista semiestructurada. Con respecto a la segunda aseercción, el análisis cualitativo demostró que la división de alumnos en grupos permitió distinguir entre distintos niveles de desempeño, sobre todo en relación con el proceso de escritura y las percepciones de los estudiantes sobre su escritura académica. En consecuencia, se puede afirmar que la PE del EXELEAA es extrapolable a las habilidades de escritura académica que se requieren en el posgrado.

CAPÍTULO 7

DISCUSIONES Y CONCLUSIONES GENERALES

7.1 DISCUSIONES GENERALES

Hoy en día, la noción de validez cubre como paraguas una amplia gama de factores relacionados no sólo con los exámenes, sino también con las interpretaciones que de ellos se derivan y del uso que se les da. A partir de la aportación de Messick (1989) de validez unitaria, se gestaron los modelos y enfoques para validar exámenes propuestos por Bachman, Mislevy y Kane. Estos enfoques han aportado copiosamente al entendimiento de la relación que existe entre un examen y el contexto del cual emerge y para el cual se toma una decisión. Todas estas relaciones se encuentran vinculadas a través de la argumentación que parte de los datos recolectados en un examen hacia las interpretaciones que de ellos establecemos sobre el futuro desempeño del candidato en un contexto dado.

El argumento de validación de la PE del EXELEAA se diseñó partiendo del enfoque de Kane (2006, 2013), pero también a partir de las propuestas que recientemente han surgido para validar un ECLAA (Chapelle, 2012; Chapelle et al., 2008, 2010). Estos autores, quienes han trabajado profusamente en el diseño del argumento de validación del TOEFL iBT, mencionan que dependiendo de las inferencias que se deseen respaldar, se debe recolectar información que permita documentar el proceso de validación. Naturalmente, la perspectiva que se adopte para diseñar el argumento de validación diferirá dependiendo de la perspectiva que se asuma; no obstante, el argumento deberá ser plausible y encontrarse bien fundamentado.

De ahí que, la investigación se haya planteado en torno a validar las inferencias semánticas que Kane (2006, 2013) propone para validar exámenes: evaluación, generalización, extrapolación y la pertinencia de las tareas con respecto al dominio para el cual se desea evaluar. La inferencia del uso que se les da a los exámenes –inferencia de tipo político– no se abordó por cuestiones de extensión de la propia investigación. Sin embargo, el uso que se le da a los exámenes es de vital importancia. De hecho, el puntaje que se le otorga a un candidato no significa nada si éste no viene acompañado de descriptores claros de lo que el examinado es capaz de hacer. De igual forma, esta información debe poder ser interpretada de forma clara y significativa por quienes toman decisiones respecto del desempeño de los sustentantes. Por

ende, las consecuencias benéficas de los exámenes deberán siempre superar las deficiencias de éstos (Bachman, 2005; Bachman y Palmer, 2010; Kane, 2013).

El primer estudio, de tipo mixto y con carácter exploratorio, se diseñó para dar cuenta de la inferencia del dominio y de la pertinencia de las tareas incluidas en la PE del EXLEEEA. Como se pudo apreciar, los géneros discursivos en el ámbito académico son diversos. Éstos han sido sistematizados de diferentes formas en el contexto anglosajón, pero poco se sabe de qué tipo de trabajos escritos les son solicitados a los estudiantes en contextos hispanohablantes. El estudio que se condujo con los docentes de posgrado permitió evidenciar que en las Ciencias Sociales y en las Humanidades y Artes, los trabajos o proyectos escritos les son frecuentemente solicitados a los estudiantes para evaluar su conocimiento. En contraste, se pudo apreciar que en las Ciencias Físico-Matemáticas e Ingenierías, así como en las Ciencias Químicas, Biológicas y de la Salud, generalmente se evalúa a través de exámenes escritos. Esta primera aproximación a un fenómeno poco estudiado: la escritura académica, nos permite percatarnos que el uso de la lengua presenta características disímiles entre áreas del saber. Si se explorara más a fondo, seguramente se podría evidenciar que incluso los géneros discursivos manifiestos en los trabajos escritos de los estudiantes varían considerablemente de una disciplina a otra.

Con respecto a las habilidades de escritura que son importantes para un desempeño académico eficiente, resulta claro que los docentes priorizan aquellas habilidades relacionadas con la construcción del conocimiento y también aquellas de tipo lingüístico; habilidades que se encuentran íntimamente relacionadas. De hecho, incluso tratándose de lengua materna, éstas últimas no pueden darse por sentadas, tal como lo demostró el estudio conducido por González et al. (2014), en el cual se evidenció que los estudiantes de recién ingreso a la universidad en la Ciudad de México carecen de habilidades lingüísticas en lengua materna; sobre todo, en las escritas. Es importante aclarar que hay habilidades académicas que quizás resulten más importantes en determinadas disciplinas que en otras, pero es cierto que los docentes del posgrado de la UNAM consideraron que la mayoría de éstas (26 de 33) eran extremadamente importantes para el éxito académico de sus alumnos de posgrado.

Las tareas de un examen con fines específicos deberían reflejar, en principio, el dominio para el cual se evalúa. Hasta cierto punto esto es cierto. No obstante, Kane et al. (1989) apuntan que en los exámenes de alto impacto debe establecerse un balance entre la autenticidad y la estandarización. En este tipo de exámenes, la estandarización es un concepto fundamental ya

que es imprescindible garantizar en todo momento que las diferentes versiones de un mismo examen son equiparables entre sí. Esto quiere decir que si un candidato sustenta la versión A, B o C, deberá obtener resultados muy similares –siempre y cuando la administración se haga en condiciones similares, la habilidad del sustentante no se haya modificado y sus características personales no hayan variado–.

Debido a la naturaleza de la escritura académica, las tareas propuestas para el ECLAA son pertinentes, mas no exhaustivas. Esto quiere decir que se debería explorar la incorporación de tareas integradas –provenientes de fuentes auditivas y escritas– para observar otras habilidades de escritura (i.e., habilidades para sintetizar, resumir y parafrasear) que no son apreciables mediante tareas discretas. Cumming (2013, 2014) afirma que la inclusión de tareas integradas conlleva diversas dificultades, mismas que no son insuperables, aunque sí requieren de cuidado. A este respecto, como se observó en los textos producidos por los estudiantes, cuando existe input escrito, los examinados tienden a copiar paquetes léxicos. Aunque Staples et al. (2013) afirman que esto tiende a disminuir en los examinados más proficientes, esto no necesariamente se evidenció así en quienes sustentaron el EXELEAA. Existen formas alternativas para mitigar este efecto. En primer lugar, se le podrían proporcionar textos a los examinados para que los leyeran y tomaran notas. Posteriormente, si se les retirara el texto fuente y se les proporcionara en ese momento la tarea de escritura, podrían hacer uso de sus notas y memoria de trabajo para resolver la tarea. Esto incentivaría también la habilidad de tomar notas para la escritura. Otra forma, aunque poco viable debido a lo costosa que resulta, sería contar con un software que permitiera calcular el porcentaje de paquetes léxicos que han sido tomados literalmente del texto. Mediante este tipo de tareas integradas, se podrían evaluar las habilidades de escritura dirigidas hacia la construcción del conocimiento.

En el caso de las tareas integradas provenientes de fuentes auditivas, aunque éstas no son prototípicas de la escritura académica (i.e., redacción de trabajos académicos), serían de gran utilidad para que los examinados demostraran habilidades escritas para resumir o sintetizar ideas principales. Una de las principales ventajas de este tipo de tareas integradas radica en que el procesamiento del input auditivo no puede ser copiado literalmente por el examinado, quien se vería obligado a hacer un ejercicio mental de procesamiento y decodificación fonética para poderlo plasmar por escrito en notas y esquemas que, posteriormente, al igual que la memoria de trabajo, le serán de utilidad para escribir sobre lo que escuchó, empleando habilidades de resumen o síntesis de ideas principales.

Finalmente, cabe señalar que tanto la comprensión auditiva como la escrita, por lo menos en el contexto académico, tienen como finalidad primordial la producción oral y/o escrita; es decir, los estudiantes normalmente desarrollan la habilidad de lectura para escribir o para interactuar de forma oral en el salón de clase (también conocidas como lectoescritura o literacidad). En este sentido, se podría prescindir de la evaluación directa de la comprensión de lectura y auditiva, y se podría integrar a la expresión oral y escrita. Esto permitirá que no se excluyeran las tareas discretas para producción escrita, sino que más bien se incluyeran más tareas escritas y de producción oral, pero de tipo integrado.

De forma alternativa, con el fin de evidenciar las habilidades destinadas a la construcción del conocimiento, también se podrían establecer otros criterios de admisión tales como presentar trabajos académicos (p.ej., capítulos de tesis, artículos, trabajos finales extensos) en español producidos durante el ciclo académico anterior. Esto permitiría, de forma más auténtica, observar cuáles son las habilidades de construcción del conocimiento con que cuenta un prospecto universitario no hispanohablante. Desafortunadamente, la desventaja de emplear dichos procedimientos de selección recae en la imposibilidad de asegurarse que, efectivamente, el trabajo escrito fue redactado por el estudiante sin ayuda de terceros. En suma, cualquier toma de decisiones conlleva ventajas y desventajas; por tal motivo, lo primordial radica en explorar qué otras opciones serían viables y sopesar los pros y contras para adoptar la decisión que proporcione el mayor número de consecuencias benéficas encaminadas a una toma de decisiones justa y equitativa.

En relación con el segundo estudio, se pudo demostrar, mediante el análisis de los resultados cuantitativos y cualitativos, que la calidad del proceso de corrección fue adecuada. Es decir, la segunda versión de la rúbrica superó la funcionalidad de la primera y resultó apropiada –aunque con ligeras modificaciones aún latentes– para medir las habilidades de escritura que requieren los examinados. Claro está, como se pudo observar, que la rúbrica no mide las habilidades relacionadas con la construcción del conocimiento, ni tampoco aquellas relacionadas con la incorporación del conocimiento previo, pero como se discutió previamente, la rúbrica es dependiente de las tareas de escritura que se seleccionan. Eventualmente, cuando se lleven a cabo las primeras revisiones de reestructuración del EXELEAA, deberá explorarse la posibilidad de incorporar tareas de escritura integradas.

En cuanto a la percepción de los evaluadores sobre la rúbrica, resaltaron los comentarios negativos narrados por la evaluadora 4, misma que demostró una evaluación deficiente y asistemática, por lo cual tuvo que ser excluida del estudio, según lo propone Linacre (2013). Fuera de estos comentarios negativos respecto de la falta de comprensión de la rúbrica y de las sesiones de retroalimentación, la percepción de los otros tres evaluadores novatos fue, en general, positiva. No obstante, destacaron los comentarios relacionados con la necesidad de contar con más ejemplos de cada uno de los niveles de desempeño de las distintas categorías contenidas en la rúbrica. De igual forma, aunque expresaron que las sesiones de capacitación y retroalimentación fueron de gran utilidad para comprender y homogeneizar los criterios de evaluación, también destacaron que les habría sido de gran ayuda tener más sesiones de capacitación. Entonces, valdría la pena contar con sesiones de capacitación grupales en las cuales poder discutir dudas respecto de ciertos aspectos que les resulten difíciles de evaluar. Eckes (2009) menciona que cada texto representa el producto de una compleja interacción que tiene lugar entre las características individuales y el conocimiento del examinado y el tema que debe desarrollar. De igual manera, este producto genera un complejo proceso cognitivo entre las características personales del evaluador y la representación mental que se éste se hace tanto del texto como de la rúbrica, para asignarle un puntaje a cada una de las categorías de la rúbrica. Por esta razón, siguiendo la línea de Knoch (2007, 2009, 2011) y de Weigle (2002), es crucial contar con un manual de capacitación para que se conduzcan sesiones de capacitación tanto para evaluadores novatos como experimentados. En suma, la rúbrica analítica per se no es suficiente para que los evaluadores puedan construirse una representación mental de la evaluación de pruebas de desempeño, como la escrita. Aun cuando los evaluadores asintieron que hizo falta más práctica y capacitación, la calidad de sus evaluaciones fue adecuada.

El análisis de MFRM representa una valiosa herramienta estadística para evaluar la calidad de las distintas facetas de variabilidad que inciden en las pruebas de respuesta construida. Desafortunadamente, el análisis no nos permite conocer la razón por la cual surgen problemas cuando éstos se presentan. Es decir, si un evaluador se encuentra fuera de los parámetros de calidad preestablecidos por el programa, no podemos saber qué es lo que está causando una evaluación deficiente. Lo mismo sucede con las facetas relacionadas con las tareas y con la rúbrica. Cuando hay alguna tarea que es muy fácil o muy difícil, no podemos deducir a partir de los resultados estadísticos qué está sucediendo con esa tarea. Lo mismo acontece con las categorías de la rúbrica. Puede ser que los niveles de desempeño estén mal redactados, que no sean claros o que simplemente no puedan evidenciarse y, por lo tanto, evaluarse en textos que

redactan los candidatos. Por esta razón, los estudios cualitativos complementarios son indispensables para comprender por qué existen fallas en el proceso de evaluación.

Finalmente, en el tercer estudio se abordó la inferencia de extrapolación. Hasta la fecha, los estudios que se han realizado en este sentido han sido contradictorios (cfr. Avdi, 2011; Cotton y Conrow, 1998; Cho y Bridgeman, 2012; Craven, 2012; Feast, 2002) y no han podido aportar información sólida que permita vincular los resultados de un ECLAA con otros instrumentos provenientes del ámbito académico. Lo primero que destacó de los resultados obtenidos de este estudio fue la alta correlación de la PE del EXELEAA con los resultados de la entrevista. Quienes obtuvieron un puntaje alto en la PE también demostraron, en la entrevista, contar con habilidades de escritura académica competentes. De igual manera, quienes puntuaron bajo en la PE escrita denotaron deficiencias y carencias en sus habilidades de escritura en el ámbito académico. Este resultado fue prometedor por varias razones. En primer lugar, las correlaciones de extrapolación en estudios previos generalmente han sido bajas, pero lo que más llama la atención es que, aun cuando la PE del EXELEAA no incluye tareas integradas que permitan dar cuenta de las habilidades destinadas a la construcción de conocimiento, la correlación entre la PE y el nivel de escritura académico evidenciado en la entrevista fue fuerte. Lo anterior sugiere que aunque la entrevista fue dirigida a dar cuenta del proceso de escritura, no se hacía referencia directa a profundizar en las habilidades con que los estudiantes cuentan para construir el conocimiento. Más aún, parece ser, según reportaron los estudiantes y docentes entrevistados en el primer estudio, la habilidad para construir el conocimiento puede transferirse de una lengua a otra. Weigle (2002) también reporta estudios similares que han encontrado estos factores de transferencias entre habilidades de una lengua a otra. En otras palabras, un estudiante que cuenta con una formación académica sólida en su lengua materna, transfiere estas habilidades de escritura académica al escribir en español. Entonces, si sus habilidades lingüísticas en español son adecuadas, su escritura académica en español se verá facilitada.

En cuanto al proceso de escritura, se evidenció que los estudiantes menos proficientes hacen un ejercicio de traducción al momento de escribir en español. Este ejercicio de traducción puede ser mental o escrito. En el primer caso, el estudiante piensa en otra lengua y escribe en español. Este proceso es ineficiente, pues va en detrimento completamente del flujo de ideas y de la saturación de la memoria de trabajo enfocada a la selección léxica y/o gramatical. Hayes (1996) y Weigle (2002) también confirmaron estas deficiencias en quienes escriben en segundas lenguas. En el segundo caso, el alumno piensa y escribe en otra lengua y, posteriormente, una

vez que ha desarrollado sus ideas por escrito, las traduce al español. La ventaja de esta aproximación es que no se interrumpe el hilo del pensamiento, pero la gran desventaja es la cantidad de tiempo que conlleva escribir y después traducir. Adicionalmente, los estudiantes que siguen este proceso de escritura son completamente dependientes del diccionario; no pueden escribir en español sin recurrir constantemente al traductor en busca de léxico o de formas gramaticales adecuadas. Adicionalmente, debido a que no cuentan con el nivel de lengua suficiente, también presentan problemas de encontrar la acepción que mejor convenga al contexto.

Contrariamente, los estudiantes con mayor dominio de escritura en español favorecen la fluidez del pensamiento escrito, ante todo. Es decir, hay quienes escriben en español e intercalan palabras en otro idioma con tal de no perder una idea, pero hay quienes prefieren ver que todo su escrito está en español, aunque tengan que recurrir eventualmente al diccionario. Parece ser que no hay una “mejor” estrategia de cómo escribir, sino personalidades distintas a quienes les sienta mejor escribir de una u otra forma. También destaca en este grupo el uso independiente, eficiente y selectivo de traductores electrónicos –todo apunta al desuso del diccionario en papel, a causa del tiempo que consume buscar palabras en él– fundamentalmente destinado al enriquecimiento léxico. Una dificultad latente, incluso en los estudiantes con nivel de escritura avanzado, es darle ese matiz académico a la escritura. Weigle (2002) se refiere a esta situación particular como la desventaja de factores sociales y culturales, independientes de los lingüísticos, a los cuales se enfrenta el no nativohablante al desconocer las expectativas del lector de otra cultura (p. 36). Esto no solamente se relaciona con el léxico especializado de la disciplina académica a la que pertenecen, sino también con expresiones y frases que son típicas de la escritura académica en español.

Otro aspecto importante que se percibió durante las entrevistas fue la alta motivación de los estudiantes no hispanohablantes del posgrado. Haciendo eco a lo que mencionan Kormos (2012), Fox (2004) y Weigle (2002) sobre la motivación y la participación del estudiante, ésta es indispensable para evitar el fracaso académico, pese a las carencias y deficiencias comunicativas en la lengua meta. En el caso de los estudiantes, sus motivaciones para estudiar en México son generalmente de tipo intrínsecas. Por ejemplo, los estudiantes no hispanohablantes que ingresan a la UNAM lo hacen por diversos motivos; uno de ellos es para hacer estudios comparativos entre sus países y México o Latinoamérica. Otra razón es por la naturaleza de la misma disciplina académica (p.ej., estudiar vulcanología o estudios

antropológicos o mesoamericanos), que es particularmente relevante en México. También es importante mencionar que gran cantidad de los estudiantes entrevistados ya residían permanentemente y laboraban en México antes de comenzar sus estudios de posgrado. Adicionalmente, una cantidad considerable de estudiantes manifestó tener una pareja sentimental mexicana. Queda claro, entonces, que todas estas motivaciones intrínsecas podrían impulsar un avance mucho más rápido de las habilidades lingüísticas de los estudiantes y mitigar significativamente del fracaso académico. En resumen, si bien es cierto que quienes se encuentran por debajo del umbral mínimo de dominio de habilidades de escritura demuestran dificultades en su proceso de escritura académica en español y relatan emociones de minusvalía de su trabajo, lo cual los pone en riesgo potencial de abandono de sus estudios o de continua exposición a la frustración al momento de escribir, no necesariamente están destinados al fracaso académico.

De hecho, esto pone de manifiesto que los criterios de selección para que un candidato ingrese a un posgrado deberían ser holísticos. Esto quiere decir que todas las formas de evaluación a considerar deberían ser incluyentes y con un enfoque siempre en las fortalezas de cada uno de los aspirantes universitarios. Por esta razón, los antecedentes académicos, la motivación del estudiante, la claridad y objetividad plasmada en un proyecto de investigación, la obtención de un buen puntaje en el EXELEAA, así como los demás instrumentos de evaluación que cada posgrado decida implementar, deberían ser ponderados como un todo para la selección de los candidatos más aptos. Queda claro que los estudiantes menos proficientes, en cuanto a sus habilidades de escritura, presentarán dificultades al escribir trabajos académicos. No obstante, si su perfil académico y su desempeño en los distintos criterios a considerar en el proceso de selección son adecuados, un acompañamiento y apoyo en el desarrollo de sus habilidades lingüísticas, como se pudo apreciar en el estudio de Fox (2004), propiciaría un óptimo desempeño académico en sus estudios de posgrado.

7.2 CONCLUSIONES GENERALES

Los nuevos enfoques argumentativos para validar exámenes han abierto una brecha en lo complejo que anteriormente representaba garantizar la funcionalidad de un instrumento de evaluación. De hecho, anteriormente, la noción de validez representaba tan solo una pequeña parte de las cualidades que debía poseer un examen. Actualmente, la validez engloba otras características de los exámenes (i.e., el uso, las consecuencias, la ética, la confiabilidad, la

fiabilidad, la practicidad y los distintos tipos de validez: de constructo, contenido, cara, predictiva). Los estudios que hoy en día se conducen en este sector se diseñan a través de un hilo argumentativo que permiten darle solidez a un examen. De ahí que, siguiendo el enfoque argumentativo para validar exámenes que Kane (2002b, 2006, 2013) y Kane et al. (1989) proponen, se diseñara el argumento de validación de la presente investigación. Con base en la evidencia proporcionada en cada una de las cuatro inferencias, y retomando la aseveración general que se estableció para esta investigación, se puede asentir que la PE del EXELEAA mide eficazmente, aunque no en toda su extensión, las habilidades de escritura que requieren los estudiantes de posgrado no hispanohablantes para ser competentes a nivel posgrado.

Como se pudo apreciar en la investigación, la descripción y conocimiento del dominio – los cuales conducen a la concepción y diseño de un ECLAA– representan un reto en el cual deberían colaborar distintos actores (i.e., docentes, coordinadores de posgrado, estudiantes y figuras institucionales en calidad de diseñar planes curriculares). El conocimiento del ámbito para el cual deseamos evaluar es imprescindible para construir exámenes que permitan tomar decisiones acertadas y con consecuencias benéficas para las distintas partes interesadas. Esto no significa que la evaluación debería convertirse en un espejo de la realidad, puesto que la autenticidad de un examen busca abstraer y poner en balance el uso de la lengua en un contexto determinado. De esta manera, se logrará diseñar exámenes de lengua estandarizados que permitan ser generalizables a un ámbito en particular; en este caso, al académico.

En cuanto a las inferencias de evaluación y generalización, es importante garantizar la calidad dentro del proceso de corrección. La concepción, diseño y revisión de una rúbrica son necesarias para contar con un instrumento que sea correctamente interpretado y utilizado por los evaluadores. En el caso de la presente investigación, la calidad dentro del proceso de corrección fue respaldada satisfactoriamente mediante los análisis estadísticos de MFRM, aunque hay que subrayar que se requiere contar con sesiones de capacitación, monitoreo, retroalimentación y una guía que favorezcan la estandarización de los criterios de corrección entre los evaluadores.

El caso de la inferencia de la extrapolación, a diferencia de los estudios que hasta la fecha se han conducido en este campo y que han demostrado bajas correlaciones entre los resultados de un ECLAA y otros instrumentos provenientes del ámbito para el cual se evalúa, la fuerte correlación entre los resultados de la PE y la entrevista arrojan luz en un inhóspito terreno dentro de la evaluación. La comparación sincrónica de las habilidades de escritura demostradas en el

examen con otras muestras de escritura provenientes del ámbito académico parecen ser una condición *sine qua non* para garantizar la plausibilidad de este tipo de estudios.

Sin embargo, los exámenes, sin importar de qué tipo, presentan limitaciones que deben ser consideradas. En primer lugar, no hay todavía ningún ECLAA que resuelva satisfactoriamente la interrogante de cómo evaluar las habilidades de escritura destinadas a la construcción del conocimiento. Por ello, sería importante explorar otras opciones como la inclusión de tareas integradas u otras formas de evidenciar que el candidato posee dicha habilidad (p.ej., mediante trabajos académicos escritos en la lengua meta). En segundo lugar, aun cuando un examinado demostrara en un ECLAA no poseer habilidades de escritura suficientes para el ámbito académico –lo cual repercutiría negativamente en su proceso de escritura académica– esto no le auguraría de ninguna manera un fracaso inminente, aunque si latente. Por ende, reconociendo la importancia de la evaluación y sus implicaciones en la vida de los examinados, se aboga por las consecuencias benéficas que debería conllevar la evaluación; es decir, por la toma de decisiones de admisión al posgrado basada en criterios holísticos y proporcionando el apoyo requerido a quienes así lo necesitan.

REFERENCIAS BIBLIOGRÁFICAS

- Alderson, J., y Clapham, C. (1993). *Examining the ELTS Test: An Account of the First Stage of the ELTS Revision Project – Research Report 2*. The British Council/University of Cambridge Local Examinations Syndicate.
- Alderson, J., Clapham, C., y Wall, D. (1995). *Language testing construction and evaluation*. Cambridge: Cambridge University Press.
- American Educational Research Association, American Psychological Association, y National Council on Measurement in Education (1974). *Standards for educational and psychological tests*. Washington, DC: American Educational Research Association.
- (1985). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- (1999). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.
- Andréu, J. (2001). Las técnicas de análisis de contenido: Una revisión actualizada. *Documento de trabajo, S2001/03*, Centro de estudios andaluces.
Recuperado del portal: <http://public.centrodeestudiosandaluces.es/pdfs/S200103.pdf>
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561–573.
- Angulo, N. (2013). La cita en la escritura académica. *Innovación Educativa*, 13 (63), 96-116.
- Attali, Y., Lewis, W., y Steier, M. (2012). Scoring with the computer: Alternative procedures for improving the reliability of holistic essay scoring. *Language Testing*, 30 (1), 125-141.
- Avdi, E. (2011). IELTS as a predictor of academic achievement in a master's program. *English Australia Journal*, 26 (2), 42-49.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19, 453-476.
- (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2, 1–34.
- (2006). Generalizability: A journey into the nature of empirical research in applied linguistics. En M. Chalhoub-Deville, C. A. Chapelle, y P. Duff (Eds.), *Inference and Generalizability in Applied Linguistics: Multiple Perspectives*. (pp. 165-208). Amsterdam, Netherlands: Benjamins.
- Bachman, L., y Palmer, A. (1996). *Language Testing in Practice*. Oxford University Press.
- (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.
- Bardin, L. (1986). *El análisis de contenido*. Madrid, Akal.
- Barkaoui, K. (2007). Rating scale impact on EFL essay marking: A mixed-method study. *Assessing Writing*, 12 (2), 86–107.

- (2010a). Variability in ESL essay rating processes: The role of the rating scale and rater experience. *Language Assessment Quarterly*, 7(May), 54–74.
- (2010b). Explaining ESL essay holistic scores: A multilevel modeling approach. *Language Testing*, 27(4), 515–535.
- (2011). Think-aloud protocols in research on essay rating: An empirical study of their veridicality and reactivity. *Language Testing*, 28(1), 51–75.
- (2014). Multifaceted Rasch Analysis for Test Evaluation. In: A. Kunnan (Ed.), *The Companion to Language assessment*. Vol. 3. Chapter 77 (pp. 1301-1322). Oxford, United Kingdom: Wiley-Blackwell.
- Bazeley, P. (2009). Analysing Qualitative Data: More than 'Identifying Themes'. *Malaysian J. Qualitative Research*, 2, 6-22.
- Bazeley, P., y Jackson, K. (2013). *Qualitative Data Analysis with NVivo*. Segunda edición. Londres: Sage.
- Becker, A. (2011). Examining Rubrics Used to Measure Writing Performance in U.S. Intensive English Programs. *The CATESOL Journal*, 22 (1), 113–130.
- Bejar, I. (2012). Rater cognition: Implications for validity. *Educational Measurement: Issues and Practice*, 31(3), 2–9.
- Beneitone, P., Esquetini, C., González, J., Marty, M., Siufi, G., y Wagenaar, R. (2007). Reflexiones y perspectivas de la Educación Superior en América Latina. Informe final. *Proyecto Tuning América Latina 2004-2007*. Bilbao: Universidad de Deusto.
- Bereiter, C., y Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum.
- Bernauer, J., Lichtman, M., Jacobs, C., y Robertson, S. (2013). Blending the Old and the New: Qualitative Data Analysis as Critical Thinking and Using NVivo with a Generic Approach. *The Qualitative Report*, 18, 1-10.
- Bond, T., y Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2a ed.). Mahwah, NJ: Erlbaum.
- Borsboom, D., Mellenbergh, G. J., y Van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111, 1061–1071.
- Borsboom, D., y Markus, K. A. (2013). Truth and Evidence in Validity Theory. *Journal of Educational Measurement*, 50 (1), 110–114.
- Breyer, F. J., Attali, Y., Williamson, D. M., Ridolfi-McCulla, L., Raminenin, C., Duchnowski, M., y Harris, A. (2014). A Study of the Use of the e-rater Scoring Engine for the Analytical Writing Measure of the GRE revised General Test. (*ETS Research Report, RR-14.24*). Princeton, NJ: ETS.
- Bryman, A. (2012). *Social Research Methods*. (4ta Edición). Nueva York: Oxford University Press.
- Canale, M. (1983). From Communicative Competence to Communicative Language Pedagogy. En J. Richards y R. Schmidt (Eds.). *Language and Communication*. Londres: Longman.
- Canale, M., y Swain, M. (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, 1 (1), 1-47.

- Canseco, G., y Byrd, P. (1989). Writing required in graduate courses in business administration. *TESOL Quarterly*, 23, 305–316.
- Carlino, P. (2004a). Escribir a través del currículum: tres modelos para hacerlo en la universidad, *Lectura y Vida. Revista latinoamericana de lectura*, 25 (1) 16-27.
- (2005b). Representaciones sobre la escritura y formas de enseñarla en universidades de América del Norte. *Revista de Educación*, 336, 143-168.
- (2007). Estudiar, escribir y aprender en universidades australianas. *Textura*, 6 (9) 11-33.
- (2013). Alfabetización académica diez años después. *Revista Mexicana de Investigación Educativa*, 18 (57) 355-381.
- Carlson, S., y Bridgeman, B. (1986). Testing ESL student writers. En K. L. Greenberg, H. S. Wiener, y R. A. Donovan (Eds.), *Writing assessment* (pp. 126–152). Nueva York: Longman.
- Castelló, M., González, D., e Iñesta, A. (2010). La regulación de la escritura académica en el doctorado: El impacto de la revisión colaborativa en los textos. *Revista Española de Pedagogía*, 247, 521-537.
- Castelló, M., e Iñesta, A. (2012). Texts as artifacts-in activity: Developing authorial identity and academic voice in writing academic research papers. En M. Castelló y C. Donahue (Eds.). *University writing: Selves and Texts in Academic Societies*. Londres: Emerald Group Publishing Limited.
- Castelló, M., Mateos, M., Castells, N., Iñesta, A., Cuevas, I., y Solé, I. (2012). Academic Writing Practices in Spanish Universities, *Electronic Journal of Research in Educational Psychology*, 10 (2), 569-590.
- Chan, S. H., Wu, R. Y., y Weir, C. J. (2014). Examining the context and cognitive validity of the GEPT Advanced Writing Task 1: A comparison with real-life academic writing tasks. *LTTC-CRELLA Collaboration Project RG-03*. Taiwan.
- Chapelle, C. (2012). Validity argument for language assessment: The framework is simple. *Language Testing*, 29 (1) 19-17.
- Chapelle, C., Enright, M., y Jamieson, J. (2008). *Building a validity argument for the test of English as a foreign language*. Nueva York: Routledge.
- (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice*, 29 (1), 3–13.
- Chitez, M., y Kruse, O. (2012). Writing cultures and genres in European higher education. En M. Castelló y C. Donahue (Eds.). *University writing: Selves and Texts in Academic Societies*. Londres: Emerald Group Publishing Limited.
- Cho, Y., y Bridgeman, B. (2012). Relationship of TOEFL iBT scores to academic performance: Some evidence from American universities. *Language Testing*, 29 (3), 421–442.
- Condon, W. (2009). Looking beyond judging and ranking: Writing assessment as a generative practice. *Assessing Writing*, 14, 141-156.
- Cooper, A., y Bikowski, D. (2007). Writing at the graduate level: what tasks do professors actually require? *Journal of English for Academic Purposes* 6 (3) 206–21.

- Cotton, F., y Conrow, F. (1998). An investigation into the predictive validity of IELTS amongst a group of international students studying at the University of Tasmania. *IELTS Research Reports Volume 1*, IELTS Australia Pty Limited, Canberra, 72–115.
- Craven, E. (2012). The quest for IELTS 7.0: Investigating English language proficiency of international students in Australian universities. *IELTS Research Reports Volume 13*, IELTS Australia, Canberra and British Council, London, 1–61.
- Creswell, J. W. (2005). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (2nd ed.). N.J.: Pearson Merrill Prentice Hall.
- Creswell, J. W., y Creswell, J. W. (2013). *Qualitative inquiry & research design: Choosing among five approaches*. Los Angeles: SAGE Publications.
- Creswell, J. W., y Plano Clark V. (2011). *Designing and Conducting Mixed Methods Research*. 2a Ed. Thousand Oaks, C.A.: Sage Publications.
- Cronbach, L. (1988). Five perspectives on validity argument. En H. Wainer y H. Braun (Eds.), *Test validity* (pp. 3–17). Hillsdale: Lawrence Erlbaum.
- Cronbach, L., y Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- Council of Europe (2001): Common European Framework of Reference for Languages: learning, teaching, assessment. Cambridge, Cambridge University Press.
- Cuatlapantzi, G., y Perales, M. (2010). How Teachers and Students at a BA in ELT Express Their Expectations About Academic Writing. En Perales (Ed.) *Literacy in Mexican Higher Education: Texts and Contexts*. Benemérita Univesidad Autónoma de Puebla. México.
- Cumming, A. (2013). Assessing Integrated Writing Tasks for Academic Purposes: Promises and Perils. *Language Assessment Quarterly*, 10 (1), 1-8.
- (2014). Assessing Integrated Skills. En A. Kunnan (Ed.), *The Companion to Language assessment. Vol. 1*. Capítulo 13 (pp. 216-229) Oxford, United Kingdom: Wiley-Blackwell.
- Cumming, A., y Berwick, R. (1996). *Validation in Language Testing. Modern Languages in Practice*. Reino Unido: Cromwell Press.
- Cumming, A., Kantor, R., y Powers, D.E. (2001). Scoring TOEFL Essays and TOEFL 2000 Prototype Writing Tasks: An Investigation Into Raters' Decision Making and Development of a Preliminary Analytic Framework. *TOEFL Report MS-22*. Princeton, Educational Testing Service.
- Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., y James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for Next Generation TOEFL. *Assessing Writing*, 10, 5–43.
- Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, 18(1), 7–24.
- Delcambre, I., y Donahue, C. (2012). Academic writing activity: Student writing in transition. En M. Castelló y C. Donahue (Eds.). *University writing: Selves and Texts in Academic Societies*. Londres: Emerald Group Publishing Limited.
- DiPardo, A., Storms, B. A., y Selland, M. (2011). Seeing voices: Assessing writerly stance in the NWP Analytic Writing Continuum. *Assessing Writing*, 16(3), 170–188.

- East, M. (2006). The impact of bilingual dictionaries on lexical sophistication and lexical accuracy in tests of L2 writing proficiency: A quantitative analysis. *Assessing Writing*, 11, 179–197.
- (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing*, 14 (2), 88–115.
- Eckes, T. (2008b). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing*, 25, 155–185.
- (2009a). Many-facet Rasch measurement. In S. Takala (Ed.), Reference supplement to the manual for relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (Section H). Strasbourg, France: Council of Europe/Language Policy Division.
Recuperado del portal de Internet: www.coe.int/t/dg4/linguistic/manuel1_EN.asp?#P19_2121
Consultado el 21 de agosto de 2014.
- (2012). Operational Rater Types in Writing Assessment: Linking Rater Cognition to Rater Behavior. *Language Assessment Quarterly*, 9(3), 270–292.
- Encinas, F., y Keranen, N. (2010). Curriculum redesign and ELT university sector writing: responding to global factors interpreted by local players. En Perales (Ed.) *Literacy in Mexican Higher Education: Texts and Contexts*. Benemérita Universidad Autónoma de Puebla. México.
- Engelhard, G. (2008b). Standard errors for performance standards based on bookmark judgments. *Rasch Measurement Transactions*, 21, 1132–1133.
- Engelhard, G., y Wind, S. A. (2013). Rating Quality Studies Using Rasch Measurement Theory. Research Report 3. New York: The College Board.
- Enright, M. K., y Quinlan, T. (2010). Complementing human judgment of essays written by English language learners with e-rater(R) scoring. *Language Testing*, 27(3), 317–334.
- Esfandiari, R., y Myford, C. M. (2013). Severity differences among self-assessors, peer-assessors, and teacher assessors rating EFL essays. *Assessing Writing*, 18(2), 111–131.
- Feast, V. (2002). The impact of IELTS scores on performance at university. *International Education Journal*, 3 (4), 70-85.
- Flores, M. A. (2014). La competencia comunicativa escrita de los estudiantes de ingeniería y la responsabilidad institucional. *Innovación Educativa*, 14 (65), 44-59.
- Fulcher, G., y Davidson, F. (2009). Test Architecture, Test Retrofit. *Language Testing* 26 (1), 123-144.
- (2007). *Language testing and assessment: An advanced resource book*. Nueva York: Routledge.
- Fox, J. (2004). Test decisions over time: tracking validity. *Language Testing*, 21 (4), 437-465.
- Gardner, S., y Nesi, H. (2012). A Classification of Genre Families in University Student Writing. *Applied Linguistics*, 34 (1) 25-52. Oxford University Press.
- González, J. (2014). Los niveles de conocimiento. El Aleph en la innovación curricular. *Innovación Educativa*, 14(65), 133-141.
- González, R. (2014). El desempeño del EXHALING y los factores asociados. En R. González (Coord.). *Habilidades Lingüísticas de los estudiantes de primer ingreso a las instituciones de educación superior del área metropolitana de la Ciudad de México*, (pp. 29-151) México: ANUIES.

- Hale, G., Taylor, C., Bridgeman, B., Carson, J., Kroll, B., y Kantor, R. (1996). A study of writing tasks assigned in academic degree programs. *Educational Testing Service Research Report 54*. Princeton: Educational Testing Service.
- Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*. Segunda edición. Londres: E. Arnold.
- Halliday, M. A. K., y Hasan, R. (1980). *Language, context and text: aspects of language in a social-semiotic perspective*. Oxford: Oxford University Press.
- Hamp-Lyons, L. (1991b). Pre-text: Task-related influences on the writer. En L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts*. Norwood, NJ: Ablex Publishing Corporation.
- (1995). Rating non-native writing: the trouble with holistic scoring. *TESOL Quarterly*, 29, (4) 759-762.
- (2002). The scope of writing assessment. *Assessing Writing*, 8, 5-16.
- (2007). Worrying about rating. *Assessing Writing*. 12, 1-9.
- (2014). Three current, interconnected concerns for writing assessment. *Assessing Writing*, 22, A1-A4.
- Hardy, J., y Römer, U. (2013). Revealing disciplinary variation in student writing: a multi-dimensional analysis of the Michigan corpus of upper-level student papers (MICUSP). *Corpora*, 8 (2), 183-207.
- Hayes, J. (1996). A new framework for understanding cognition and affect in writing. En M. Levy y S. Ransdell, (Eds.) *The science of writing: Theories, methods, individual differences, and applications*. Mahwah, NJ: Lawrence Erlbaum, 1–27.
- Hayes, J., y Flower, L. (1980). Identifying the organization of writing processes. En L. Gregg y E. Steinberg, (Eds.) *Cognitive processes in writing: An interdisciplinary approach*. (pp. 3-30). Hillsdale, NJ: Lawrence Erlbaum.
- He, L., y Shi, L. (2012). Topical Knowledge and ESL writing, *Language Testing*, 29 (3), 443-464.
- He, T., Gou, W. J., Chien, Y., Chen, I. J., y Chang, S. (2013). Multi-Faceted Rasch Measurement and bias patterns in EFL writing performance assessment. *Psychological Reports*, 112 (2), 469–485.
- Hernández, F., y Castelló, M. (2014). Análisis de los géneros de escritura académica de acuerdo con la percepción de los estudiantes universitarios. *Innovación Educativa*, 14 (65), 61-80.
- Hernández, R., Fernández, C., y Baptista, P. (2010). *Metodología de la investigación*. México: McGraw-Hill Educación.
- Horkay, N., Bennett, R.E., Allen, N., Kaplan, B., y Yan, F. (2006). Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *The Journal of Technology, Learning & Assessment*, 5, 4–49.
- Horowitz, D. (1986). What professors actually require of students: academic tasks for the ESL classroom. *TESOL Quarterly*, 20 (3) 445-462.
- Huang, J., y Foote, C. (2010). Grading Between the Lines: What Really Impacts Professors' Holistic Evaluation of ESL Graduate Student Writing? *Language Assessment Quarterly*, 7(1), 37–41.

- Humphreys, P., Haugh, M., Fenton-Smith, B., Lobo, A., Michael, R., y Walkinshaw, I. (2012). Tracking international students' English proficiency over the first semester of undergraduate study. *IELTS Research Reports Online Series 1*.
- Huot, B., O'Neill, P., y Moore, P. (2010). A Usable Past for Writing Assessment. *National Council of Teachers of English*, 72 (5), 495-517.
- Hyland, K. (2008). Disciplinary voices: Interactions in research writing. *English Text Construction*, 1 (1), 5–22.
- Hyland, T. (2009). Drawing a line in the sand: Identifying the border zone between self and other in EL1 and EL2 citation practices. *Assessing Writing*, 14 (1), 62-74.
- Hymes, D. H. (1972). On communicative competence. En J.B. Pride and J. Holmes (Eds.), *Sociolinguistics: Selected readings*. (pp. 269-293). Harmondsworth, Middlesex: Penguin.
- Ingram, D. y Bayliss, A. (2007). IELTS as a predictor of academic language performance, Part 1', *IELTS Research Reports Volume 7*, IELTS Australia Pty and British Council, Canberra, 137–199.
- Iñesta, A., y Castelló, M. (2012). Towards an integrative unit of analysis: Regulation Episodes in Expert Research Article Writing. En C. Bazermann, C. Dean, J. Early, K. Lunsford, S. Null, P. Rogers y A. Stansell (Eds.). *International Advances in Writing Research: Cultures, Places, Measures*, (pp. 421-448).
Recuperado del portal de Internet: <http://wac.colostate.edu/books/wrab2011/chapter24.pdf>
- Jamieson, S. (2004). Likert scales: how to (ab) use them. *Medical education*, 38 (12), 1217-1218.
- Johnson, M., Mercado, L., y Acevedo, A. (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity, *Journal of Second Language Writing*, 21, 264–282.
- Jonsson, A., y Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2 (2), 130–144.
- Jurado, M. (2014). Examen de Habilidades Lingüísticas (EXHALING). En R. González (Coord.). *Habilidades Lingüísticas de los estudiantes de primer ingreso a las instituciones de educación superior del área metropolitana de la Ciudad de México*, (pp. 155-188) México: ANUIES.
- Kane, M. (1992). An argument-based approach to validation. *Psychological Bulletin*, 112, 527–535.
- (2002b). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*, 21(1), 31–41.
- (2006). Validation. En R. Brennan (Ed.). *Educational measurement*, (pp. 17–64). Westport: American Council on Education and Praeger.
- (2013). Validating the Interpretations and Uses of Test Scores, *Journal of Educational Measurement*, 50 (1), 1-73.
- Kane, M., Crooks, T., y Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice*, 18(2), 5–17.
- Kellogg, R. (1996). A model of working memory in writing. En Levy, M. y Ransdell, S. (Eds.) *The Science of writing: Theories, methods, individual differences, and applications*. Mahwah, NJ: Lawrence Erlbaum Associates, 57-72.

- Knobel, M., y Lankshear, C. (2001). *Maneras de Ver: El Análisis de Datos en Investigación Cualitativa*. Morelia: Instituto Michoacano de Ciencias de la Educación.
- Knoch, U. (2007). Diagnostic Writing Assessment. The Development and Validation a Rating Scale. Unpublished PhD, University of Auckland.
- (2007b). Do empirically developed rating scales function differently to conventional rating scales for academic writing? *Spain Fellow Working Papers in Second or Foreign Language Assessment*, 5, 1–36.
- (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26 (2), 275–304.
- (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from? *Assessing Writing*, 16, 81-96.
- Knoch,U. y Elder, C. (2010). Validity and fairness implications of varying time conditions on a diagnostic test of academic English writing proficiency. *System*, 38, 63–74.
- Kobrin, J., Deng, H., y Shaw, E. (2011). The association between SAT prompt characteristics, response features, and essay scores. *Assessing Writing*, 16, 154–169.
- Kokhan, K. (2012). Investigating the possibility of using TOEFL scores for university ESL decision-making: Placement trends and effect of time lag. *Language Testing*, 29(2), 291-308.
- (2013). An argument against using standardized test scores for placement of international undergraduate students in English as a Second Language (ESL) courses. *Language Testing*, 30(4), 467-489.
- Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, 21, 390-403.
- Kunnan, A. (2003). Test Fairness. En M. Milanovic y C. Weir (Eds.), *Select papers from the European Year of Languages Conference, Barcelona*. Cambridge: Cambridge University Press.
- Lado, R. (1961). *Language Testing: The Construction and Use of Foreign Language Tests*. Longman, London.
- Lai, E. R., Wolfe, E. W., y Vickers, D. (2015). Differentiation of Illusory and True Halo in Writing Scores. *Educational and Psychological Measurement*, 75, 102-125.
- Lee, H. (2008). The relationship between writers' perceptions and their performance on a field-specific writing test, *Assessing Writing*, 13, 93–110.
- Lee, H., y Anderson, C. (2007). Validity and topic generality of a writing performance test. *Language Testing*, 24, 307–330.
- Lee, Y.W., Gentile, C., y Kantor, R. (2008). Analytic Scoring of TOEFL CBT Essays: Scores From Humans and E-rater. *Research Reports, RR-81*, ETS, Princeton, NJ, 1–84.
- Leki, I., Cumming, A., y Silva, T. (2008). *A Synthesis of Research on Second Language Writing in English*. Routledge: New York.
- Lim, G. S. (2011). The development and maintenance of rating quality in performance writing assessment: A longitudinal study of new and experienced raters. *Language Testing*, 28(4), 543–560.

- Linacre, J. M. (1989-2015). *Facets computer program for many-facet Rasch measurement*, version 3.71.4. Beaverton, Oregon: Winsteps.com
- (2012a). Facets Tutorial 1. 1-32.
Recuperado del portal de Internet: <http://www.winsteps.com/a/ftutorial1.pdf>
- (2012b) Facets Tutorial 2. 1-40.
Recuperado del portal de Internet: <http://www.winsteps.com/a/ftutorial2.pdf>
- (2012c) Facets Tutorial 3. 1-29.
Recuperado del portal de Internet: <http://www.winsteps.com/a/ftutorial3.pdf>
- (2012d) Facets Tutorial 3. 1-18.
Recuperado del portal de Internet: <http://www.winsteps.com/a/ftutorial4.pdf>
- (2013). A User's Guide to FACETS. Rasch-Model Computer Programs. Program Manual 3.71.0.
Recuperado el 27 de septiembre de 2015 del portal: www.winsteps.com
- Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the raters? *Language Testing*, 19 (3), 246–276.
- (2005). *Assessing second language writing. The rater's perspective*. Frankfurt: Peter Lang.
- Lynch, B. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18, 351-372.
- Martin, J. R. (1997). Analysing genre: functional parameters. En F. Christie y J. R. Martin (Eds.). *Genre and institutions*. London: Cassell.
- Martin, J. R, y Rose, D. (2003). *Working with Discourse: Meaning beyond the clause*. London: Continuum
- (2008). *Genre Relations: Mapping Culture*. London: Equinox.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.
- McCurry, D. (2010). Can machine scoring deal with broad and open writing tests as well as human readers? *Assessing Writing*, 15 (2), 118–129.
- McNamara, T. (1996). *Measuring Second Language Performance*, Reino Unido: Blackwell.
- McNamara, T., y Roever, C. (2006). *Language testing: The social dimension*. Reino Unido: Blackwell.
- Mendoza, A. (2014). Las prácticas de evaluación docente y las habilidades de escritura requeridas en el nivel posgrado. *Innovación Educativa*, vol. 14, 66 (3), 147-175.
- Messick, S. (1989). Validity. En R. Linn (Ed.), *Educational measurement* (pp. 13-103). Nueva York: American Council on Education and Macmillan.
- (1995). Validity of Psychological Assessment, *American Psychologist*, 50 (9), 741-749.
- Mislevy, R. (2006). Cognitive psychology and educational assessment. En R. Brennan (Ed.), *Educational measurement* (pp. 257–305). Westport, CT: American Council on Education and Praeger.
- (2009). Validity from the perspective of model-based reasoning. En R. Lissitz (Ed.), *The concept of validity* (pp. 83–108). Charlotte, NC: Information Age Publishers.

- (2013). Modeling Language for Assessment, En C. Chapelle (Ed.), *Encyclopedia of Applied Linguistics* (pp. 1-9). Blackwell.
- Mislevy, R., Steinberg, L., y Almond, R. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3–62.
- Mogey, N., Paterson, J., Burk, J., y Purcell, M. (2010). Typing compared with handwriting for essay examinations at university: Letting the students choose. *ALT-J Research in Learning Technology*, 18 (1), 29-47.
- Mogey, N., y Hartley, J. (2013). To write or to type? The effects of handwriting and word-processing on the written style of examination essays. *Innovations in Education and Teaching International*, 50 (1), 85-93.
- Montee, M., y Malone, M. (2014). Writing Scoring Criteria and Score Reports. En A. Kunnan (Ed.), *The Companion to Language assessment*. Vol. 2. Capítulo 51 (pp. 847-859). Oxford, United Kingdom: Wiley-Blackwell.
- Moore, T., y Morton, J. (2005). Dimensions of difference: a comparison of university writing and IELTS writing. *Journal of English for Academic Purposes*, 4, 43-66.
- Myford, C. M. (2012). Rater cognition research: Some possible directions for the future. *Educational Measurement: Issues and Practice*, 31(3), 48–49.
- Myford, C. M. y Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4, 386–422.
- (2004). Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement*, 5, 189–227.
- Ong, J. (2013). Discovery of ideas in second language writing task environment. *System*, 41, 529-542
- Paul, A. (2004). IELTS as a predictor of academic language performance, Part 2. *IELTS Research Reports*, Vol. 7, 1-35.
- Petersen, J. (2009). “This test makes no freaking sense”: Criticism, confusion, and frustration in timed writing. *Assessing Writing*, 14, 178–193.
- Plakans, L. (2010). Independent vs. integrated writing tasks: A comparison of task representation. *TESOL Quarterly*, 44 (1), 185–194.
- Plakans, L., y Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing*, 22, 217–230.
- Powers, D. (2015). The incremental contribution of TOEIC Listening, Reading, Speaking, and Writing tests to predicting performance on real-life English language tasks. *Language Testing*, 32 (2), 151-167.
- Powers, D.; Kim, H., y Weng, V. (2010). The Redesigned TOEIC Listening and Reading Test: Relations to Test Taker Perceptions of Proficiency in English. *Compendium Study*. Princeton, Educational Testing Service.
- Prieto, G. (2011). Evaluación de la ejecución mediante el modelo Many-Facet Rasch Measurement. *Psicothema*, 23 (2), 233–238.

- Prieto, G. y Nieto, E. (2014). Analysis of rater severity on written expression exam using Many-Faceted Rasch Measurement. *Psicológica*, 35, 285-397.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: Mesa Press.
- Rezaei, A. R., y Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15 (1), 18–39.
- Rose, D., y Sullivan, O. (1996). *Introducing Data Analysis for Social Scientists*, 2a edición. Buckingham: Open University Press.
- Rosenfeld, M., Leung, S., y Oltman, P. (2001). Identifying the reading, writing, speaking, and listening tasks important for academic success at the undergraduate and graduate levels. *TOEFL Monograph Series MS-21*. Princeton, Educational Testing Service.
- Rosenfeld, M., Courtney, R., y Fowles, M. (2004). Identifying the Writing Tasks Important for Academic Success at the Undergraduate and Graduate Levels. *Research Report 42*. Princeton: Educational Testing Service.
- Samuel, M., y Okey, L. E. (2015). The relevance and significance of correlation in social science research. *International Journal of Sociology and Anthropology Research*, 1 (3), 22-28.
- Saxton, E., Belanger, S., y Becker, W. (2012). The Critical Thinking Analytic Rubric (CTAR): Investigating intra-rater and inter-rater reliability of a scoring mechanism for critical thinking performance assessments. *Assessing Writing*, 17 (4), 251–270.
- Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing*, 25 (4), 465-493.
- Shah, P., y Freedman, E. (2009). Bar and Line Graph Comprehension: An Interaction of Top-Down and Bottom-Up Processes, *Topics in Cognitive Science*, 1-19.
- Sinharay, S, Powers, D., Feng, Y., Saldivia, L., Giunta, A., Simpson, A., y Weng, V. (2009). Appropriateness of the TOEIC Bridge test for students in three countries of South America. *Language Testing*, 26 (4) 589-619.
- Spolsky, B. (1981). Some ethical questions about language testing. En C. Klein-Braley y D. Stevenson (Eds.), *Practice and problems in language testing* (pp. 5-21). Frankfurt, Alemania: Peter D. Lang.
- Srnka, K. J., y Koeszegi, S. T. (2007): From Words to Numbers - How to Transform Rich Qualitative Data into Meaningful Quantitative Results: Guidelines and Exemplary Study, *Schmalenbach's Business Review*, 59, 29-57.
- Staples, S., Egbert, J., Biber, D., y McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, 12, 214–225.
- Suzuki, Y. (2015). Self-assessment of Japanese as a second language: The role of experiences in the naturalistic acquisition. *Language Testing*, 32 (1) 63-81.
- Teberosky, A. (2007). El texto académico. En M. Castelló (Coord.). *Escribir y comunicarse en contextos científicos y académicos. Conocimientos y estrategias*. Barcelona: Graó.
- Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27 (2), 237-246.

- Toulmin, S. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.
- Toulmin, S., Rieke, T., y Janik, A. (1984). *An introduction to reasoning*. Nueva York: Macmillan.
- Ushioda, E., y Harsch, C. (2011). Addressing the needs of international students with academic writing difficulties: *Pilot project 2010/11, Strand 2: Examining the predictive validity of IELTS scores. Internal report*, Centre for Applied Linguistics: University of Warwick.
- Uysal, H. (2010). A critical review of the IELTS writing Test. *ELT Journal*, 64 (3), 314-320.
- van Dijk, T. A. (1985). *Handbook of Discourse Analysis*. London: Academic Press.
- van Dijk, T. A., y Kintsch, W. (1983). *Strategies of discourse comprehension*. NY: Academic Press.
- Vidal, C., y Perales, M. (2010). Writing Tasks in a Pre-Service English Teaching Program: Professors' Beliefs. En Perales (Ed.) *Literacy in Mexican Higher Education: Texts and Contexts*. Benemérita Universidad Autónoma de Puebla. México.
- Wächter, B., y Maiworm, F. (2008). English-Taught Programmes in European Higher Education: the Picture in 2007, *ACA Papers on International Cooperation in Education*. Bonn: Lemmens.
- Wang, Z., y Yaho, L. (2013). The Effects of Rater Severity and Rater Distribution on Examinees' Ability Estimation for Constructed-Response Items, *ETS Research Report No. RR-13-23*.
- Weigle, S. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- (2010). Validation of automated scores of TOEFL iBT tasks against non-test indicators of writing ability. *Language Testing*, 27(3), 335–353.
- (2015). Linguistic dimensions of impromptu test essays compared with successful student disciplinary writing: Effects of language background, topic, and L2 proficiency. *Journal of English for Academic Purposes*, 18, 25-39.
- Weigle, S., y Parker, K. (2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing*, 21, 118-133.
- Weir, C. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Basingstoke: Palgrave Macmillan.
- White, E.M. (1984). Holisticism. *College Composition and Communication*, 35 (4), 400-409.
- Widdowson, H. (2001). Communicative language testing: The art of the possible. En C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, K. O'Loughlin (Eds.), *Studies in Language Testing 11, experimenting with Uncertainty: Essays in honor Alan Davies*. Cambridge University Press.
- Wind, S. A., y Engelhard, G. (2013). How invariant and accurate are domain ratings in writing assessment? *Assessing Writing*, 18 (4), 278–299.
- Wolfe, E. W. (1997). The relationship between essay reading style and scoring proficiency in a psychometric scoring system. *Assessing Writing*, 4, 83–106.
- (2004). Identifying rater effects using latent trait models. *Psychology Science*, 46(1), 35–51.
- Wolfe, E. W., y Mcvay, A. (2012). Application of latent trait models to identifying substantively interesting raters. *Educational Measurement: Issues and Practice*, 31 (3), 31–37.

- Woodrow, L. (2006). Academic success of international postgraduate education students and the role of English proficiency. *University of Sydney Papers in TESOL*, 1, 51–70.
- Yang, H. (2012). The language learning potential of form-focused feedback on writing. *English for Specific Purposes*, 31, 174–187.
- Yao, L. (2010). *BMIRTII: Bayesian Multivariate Item Response Theory [Programa computacional]*. Defense Manpower Data Center, Monterey, CA.
- Zhang, Y., y Wildemuth, B. (2009). Unstructured interviews. En B. Wildemuth (Ed.), *Applications of Social Research Methods to Questions in Information and Library Science* (pp. 222-231). Westport, CT: Libraries Unlimited.
- (2009). Qualitative analysis of content. En B. Wildemuth (Ed.), *Applications of Social Research Methods to Questions in Information and Library Science* (pp. 308-319). Westport, CT: Libraries Unlimited.
- Zhao, C. G. (2012). Measuring authorial voice strength in L2 argumentative writing: The development and validation of an analytic rubric. *Language Testing*, 30 (2), 201-230.
- Zhizhko, E. A. (2014). La enseñanza de la escritura y lectura de textos académicos a los futuros investigadores educativos. *Innovación Educativa*, 14 (65), 99-113.

ANEXOS ESTUDIO 1

Anexo 1.1

Docentes de posgrado participantes en el estudio sobre
las habilidades de escritura requeridas en el posgrado

Área 1: Ciencias Físico-Matemáticas y de las Ingenierías	
POSGRADO	MATERIA IMPARTIDA
Ingeniería en sistemas computacionales	Mejora de calidad
	Autómatas y Lenguajes Formales
	Ingeniería de software orientada a objetos
Ingeniería	Diseño en Ingeniería Mecánica
Ciencias Físicas	Electromagnetismo
Ciencia e Ingeniería de Materiales	Procesos Cerámicos
	Biomateriales
Astrofísica	Astrofísica Relativista
	Materia Interestelar
	Procesos Radiactivos en Astrofísica
Ciencias de la Tierra	SIN DATO

Área 2: Ciencias Biológicas, Químicas y de la Salud	
POSGRADO	MATERIA IMPARTIDA
Neurobiología	Estructura y función del hipotálamo
	Estadística
Ciencias Biomédicas	Inmunología avanzada
Ciencias Médicas, Odontológicas y de la Salud	Seminario de tesis
	Seminario de investigación
Facultad de Psicología	Seminario de investigación
	Medición y evaluación
Ciencia Bioquímica	Espectroscopía de fluorescencia
	Fisiología microbiana
Ciencias Biológicas	Biología Molecular
	Tópicos Selectos de Ultraestructura aplicada a la ciencias biológicas y de la salud
Ciencias del Mar y Limnología	Oceanografía Física

Área 3: Ciencias Sociales	
POSGRADO	MATERIA IMPARTIDA
Ciencias de la Administración	Seminario de Investigación de Mercados
	Seminario de Ética y Responsabilidad Social Corporativa
	Antropología de las empresas
	Dirección de Recursos Humanos
Ciencias Políticas y Sociales	Seminario de Investigación
	Educación, apropiación y resistencia cultural
Geografía	Geografía de la Población y Medio Ambiente
Derecho	Teorías de la Argumentación Jurídica
Antropología	Bioantropología de poblaciones Contemporáneas
	Diseño de proyectos de investigación
	Teoría de la antropología física

Área 4: Humanidades y Artes	
POSGRADO	MATERIA IMPARTIDA
Letras	Seminario de cuento
	Seminario Estudios Culturales
	Mito y rito: Herramientas para la investigación literaria y teatral
	Seminario de tesis
	Ensayo. Optativa
	Ciencia ficción: temas y motivos
Estudios Mesoamericanos	Historia Colonial
	Códices mesoamericanos I
	Herramientas de Análisis Lingüístico para el procesamiento de textos orales
	Taller de redacción, primer semestre de maestría
Historia del arte	Historiografía
	Seminario de Teoría y Crítica de Arquitectura
	Tesis
	Historia del Arte Mexicano Siglo XIX. Conciencia nacionalista.
	Representación indígena en el cine latinoamericano
Arte y Diseño	Seminario de derechos de autor
	Libros de artista y publicaciones de arte en Estados Unidos y Europa 1960-2010
Arquitectura	Arquitectura contemporánea en su marco socio cultural
	Tecnología
Lingüística Aplicada	Trastornos del lenguaje
	Seminario de Investigación
	La teoría de las Representaciones Sociales y la lingüística Aplicada
Música	Filosofía de la Educación Musical
Bibliotecología y Estudios de la Información	Seminario de Comunidades Científicas y de Información
Pedagogía	Política pública en educación
Docencia para la Educación Media Superior	Diseño de materiales didácticos

Anexo 1.2

Cuestionario sobre la evaluación en el posgrado

EVALUACIÓN Y ESCRITURA ACADÉMICA EN EL POSGRADO DE LA UNAM

El siguiente cuestionario busca identificar la importancia de la escritura en los diversos programas del Posgrado de la UNAM. La información que se recabará es confidencial y usted podrá solicitar que se le envíen los resultados del estudio. El cuestionario consta de dos partes: en la primera, aparece una serie de preguntas abiertas sobre cómo evalúan los docentes a sus estudiantes, mientras que en la segunda, se le pregunta sobre la importancia de las habilidades de escritura que requieren los estudiantes que acaban de iniciar sus estudios de maestría.

De antemano, agradecemos su valiosa participación.

LA EVALUACIÓN EN EL POSGRADO

1. ¿En qué posgrado imparte cursos? (Si es en varios, elija sólo uno).

2. ¿Cuál fue la última materia que impartió y a qué semestre del plan de estudios corresponde? (Si fueron varias, elija sólo una).

3. ¿Cómo evaluó a sus alumnos y qué peso de la evaluación total le otorgó a cada sección?

Ejemplo: Asistencia y participación 20%, Examen parcial 30%, Reseña de un libro 10% y Presentación oral de un tema previamente asignado 40%

4. En caso de haber evaluado con alguna tarea/examen/trabajo escrito, describalo brevemente y mencione el número de cuartillas o palabras solicitadas en cada una de ellos.

Ejemplo: El examen parcial se resolvió en clase y constó de 10 preguntas abiertas a desarrollar (por lo menos media cuartilla cada pregunta). En la reseña, se le pidió al alumno un texto de mínimo 1000 palabras.

5. ¿Cuáles fueron sus criterios para evaluar los trabajos, tareas o exámenes escritos de sus estudiantes?

Puede elegir más de una opción

- Relevancia del contenido
- Uso apropiado de fuentes
- Capacidad de reflexión
- Coherencia y cohesión
- Precisión gramatical
- Presentación y formato de acuerdo al género discursivo
- Other:

6. Si ha tenido estudiantes de posgrado no hispanohablantes y le gustaría participar en una entrevista respecto de las dificultades que dichos estudiantes encaran en cuanto a la escritura, por favor escriba su nombre completo y correo electrónico.

Anexo 1.3

Encuesta sobre las habilidades de escritura académica en el posgrado

EVALUACIÓN Y ESCRITURA ACADÉMICA EN EL POSGRADO DE LA UNAM

*** Required**

LAS HABILIDADES DE ESCRITURA EN EL POSGRADO

Por favor evalúe del 0 al 5 cada una de las siguientes habilidades de escritura que considera son importantes para aquellos estudiantes que comienzan sus estudios de maestría en el posgrado al cual usted pertenece.

(0 no es necesaria, 1 poco importante, 2 moderadamente importante, 3 importante, 4 muy importante, 5 extremadamente importante). Traducido y adaptado de Rosenfeld et al. (2004).

1. Describir observaciones.*
por ejemplo: de un evento, comportamiento, lugar, objeto o experimento.

0 1 2 3 4 5

No es necesaria Extremadamente importante

2. Explicar cómo se lleva a cabo un procedimiento.*
por ejemplo: instrucciones para materiales o manuales

0 1 2 3 4 5

No es necesaria Extremadamente importante

3. Abstractar o resumir información esencial.*
por ejemplo: en discursos, observaciones o textos.

0 1 2 3 4 5

No es necesaria Extremadamente importante

4. Expresar opiniones personales.*
por ejemplo: con respecto a temas, situaciones o problemas.

0 1 2 3 4 5

No es necesaria Extremadamente importante

5. Explicar un hecho o suceso utilizando evidencia.*
por ejemplo: evidencia proveniente de sucesos históricos, datos o resultados de investigación.

0 1 2 3 4 5

No es necesaria Extremadamente importante

6. Analizar y sintetizar información de múltiples fuentes (incluye comparación y contraste).*

0 1 2 3 4 5

No es necesaria Extremadamente importante

7. Predecir consecuencias o resultados mediante el análisis de información, patrones o procesos.*

0 1 2 3 4 5

No es necesaria Extremadamente importante

8. Escribir persuasivamente mediante la construcción de un argumento bien razonado para apoyar o refutar una postura.*

0 1 2 3 4 5

No es necesaria Extremadamente importante

9. Explorar las relaciones entre ideas complejas y posiblemente contradictorias. *

0 1 2 3 4 5

No es necesaria Extremadamente importante**10. Examinar el razonamiento en un argumento dado y discutir las fortalezas y debilidades de su lógica. ***

por ejemplo: en la legitimidad de sus aserciones, la solidez de sus hipótesis, la suficiencia de su respaldo o la distinción entre correlación y causalidad.

0 1 2 3 4 5

No es necesaria Extremadamente importante**11. Identificar los problemas en una línea de acción propuesta o en la interpretación de hechos y proponer soluciones o alternativas de interpretación. ***

0 1 2 3 4 5

No es necesaria Extremadamente importante**12. Interpretar los datos dentro de un marco determinado. ***

por ejemplo: mediante la aplicación de hallazgos a nuevas situaciones, formulando preguntas perspicaces, identificando la necesidad de profundizar en la información u obteniendo conclusiones.

0 1 2 3 4 5

No es necesaria Extremadamente importante**13. Clasificar la información de acuerdo con categorías o jerarquías. ***

por ejemplo: en esquemas o diagramas de organización.

0 1 2 3 4 5

No es necesaria Extremadamente importante**14. Utilizar las convenciones de un género en particular. ***

por ejemplo: un ensayo, una monografía, un reporte, un abstract, etc.

0 1 2 3 4 5

No es necesaria Extremadamente importante**15. Escribir adecuadamente para un público bien informado y reflexivo. ***

por ejemplo: mantener un tono apropiado, proporcionar contexto suficiente u otra información que les permita a los lectores comprender los puntos establecidos.

0 1 2 3 4 5

No es necesaria Extremadamente importante**16. Presentar los datos y la información de manera clara y lógica. ***

por ejemplo: ofreciendo explicaciones que hagan que el material sea comprensible para una audiencia particular (incluye tablas y gráficos, así como texto).

0 1 2 3 4 5

No es necesaria Extremadamente importante**17. Usar la analogía, metáfora o comparación para definir o explicar conceptos técnicos o abstractos para una audiencia general. ***

0 1 2 3 4 5

No es necesaria Extremadamente importante

18. Utilizar formatos eficientes y claros para organizar la información y guiar al lector. *
por ejemplo: diagramas de flujo, viñetas, títulos.

0 1 2 3 4 5

No es necesaria Extremadamente importante

19. Integrar el material citado y referenciado debidamente dentro del texto. *

0 1 2 3 4 5

No es necesaria Extremadamente importante

20. Dar crédito adecuado a las fuentes. *

por ejemplo: mediante citas y notas al pie de página o al final del texto.

0 1 2 3 4 5

No es necesaria Extremadamente importante

21. Desarrollar un debate bien enfocado y debidamente sustentado mediante ejemplos relevantes. *

0 1 2 3 4 5

No es necesaria Extremadamente importante

22. Aclarar las relaciones entre ideas principales y secundarias *

0 1 2 3 4 5

No es necesaria Extremadamente importante

23. Organizar las ideas y la información coherentemente. *

0 1 2 3 4 5

No es necesaria Extremadamente importante

24. Escribir de forma precisa y concisa, evitando frases vagas o vacías. *

0 1 2 3 4 5

No es necesaria Extremadamente importante

25. Escribir con claridad mediante transiciones suaves de un pensamiento a otro. *

0 1 2 3 4 5

No es necesaria Extremadamente importante

26. Escribir con fluidez, evitando que el lenguaje sea complicado y rebuscado. *

0 1 2 3 4 5

No es necesaria Extremadamente importante

27. Variar la estructura de las oraciones para comunicar ideas de manera efectiva. *

0 1 2 3 4 5

No es necesaria Extremadamente importante

28. Expresar ideas en forma original o novedosa para mantener el interés del lector. *

0 1 2 3 4 5

No es necesaria Extremadamente importante

29. Revisar y editar el texto para mejorar su calidad, coherencia y precisión. *

0 1 2 3 4 5

No es necesaria Extremadamente importante

30. Usar vocabulario técnico y específico al contenido del tema con precisión y de forma apropiada para un propósito y audiencia en particular. *

0 1 2 3 4 5

No es necesaria Extremadamente importante

31. Elegir palabras efectivamente. *

0 1 2 3 4 5

No es necesaria Extremadamente importante

32. Utilizar correctamente la gramática del español escrito con el fin de evitar que el lector se distraiga o que se cometan errores que alteren el significado. *

0 1 2 3 4 5

No es necesaria Extremadamente importante

33. Evitar errores ortográficos. *

por ejemplo: la ortografía y la puntuación.

0 1 2 3 4 5

No es necesaria Extremadamente importante

¡Muchas gracias por su colaboración!

Agradecemos cualquier comentario adicional que desee agregar.

Anexo 1.4

Resultados del cuestionario sobre las prácticas de evaluación en el posgrado

Área 1: Ciencias Físico-Matemáticas y de las Ingenierías

3. ¿Cómo evaluó a sus alumnos y qué peso total le otorgó a cada sección?
 4. En caso de haber evaluado con alguna tarea/examen/trabajo escrito, descríballo brevemente y mencione el número de cuartillas o palabras solicitadas en cada uno de ellos.

PRÁCTICAS DE EVALUACIÓN DOCENTE	ESCRITURA
50% Tareas (3) 50% Examen	Tarea 1: Cuatro problemas que necesitan algunos cálculos, analíticos y aritméticos, y por lo menos un párrafo de interpretación. Una pregunta para explicar tres conceptos clave. Tarea 2: Dos problemas que requieren muy pocos cálculos, y la elaboración de programas para realizar los cálculos necesarios y gráficas. Un cuestionario de ocho preguntas sobre una lectura. Tarea 3: Cinco problemas que requieren tanto cálculos como interpretaciones. Examen: Se plantearon 8 preguntas (1.25pts) y 4 problemas (2.5pts). El estudiante podía elegir a qué responder sabiendo que el examen sería calificado sobre 10 puntos y con la obligación de responder por lo menos a dos preguntas y a un problema. Las respuestas a las preguntas necesitaban de aproximadamente media cuartilla. Para los problemas, depende del método elegido y de la cantidad de detalle que el estudiante incluya.
Exámenes 60% Trabajo final 30% Otras actividades 10%	Uno de los exámenes parciales constó de la resolución de un estudio de caso en grupos de 3 personas, con datos técnicos y 8 preguntas relacionadas. No se especificó un mínimo de cuartillas, pero me entregaron entre 10 y 25 cuartillas.
Tareas: 15% Exámenes parciales: 85%	Los exámenes parciales cubren cada uno aproximadamente 25% del contenido del curso. Las respuestas consisten en fórmulas matemáticas o diagramas acompañados por una breve explicación, comentario o demostración (es decir, argumentos lógicos en favor de la respuesta). Un examen típico tiene cuatro preguntas y las respuestas abarcan una o dos páginas manuscritas.
Examen final escrito 40% Examen final oral 60%	Hubo alrededor de 8 tareas que consistían en tres problemas. El lenguaje de los trabajos es principalmente matemático. El examen final escrito es similar. Hubo un tema, escogido por los alumnos de entre tres opciones a desarrollar en una cuartilla solamente. Sólo sirvió para redondear los resultados.
Tareas 70% Exámenes 30%	Examen con 10 preguntas a escoger 4. Principalmente desarrollos matemáticos.
Tareas: 20% Participación en clase: 20% Entregables de proyecto: prototipo, reporte final, presentación final: 60%	Varias de las tareas consisten en ensayos, críticas o comentarios de una cuartilla sobre lecturas. Las lecturas se refieren a capítulos del libro de texto o artículos). Los reportes de proyecto son documentos escritos cuyo contenido básico es especificado por el profesor, pero los alumnos tienen la libertad de definir su propio formato y extensión. Estos reportes van de 20 a 40 páginas por lo general.

Participación en clase 25% Exposiciones de trabajos variados hechos por ellos, escritos y verbales 75%	No en especial.
Asistencia 20% Presentación oral 50% Tareas 30%	Desarrollo de temas de investigación relacionados con el curso (10 páginas)
Asistencia y participación 20% Participación en proyecto del curso 80%	Durante el proyecto los alumnos entregan resúmenes de avances a través de documentos con gráficas y su interpretación
Participación 20% Presentación oral de un tema 20% Exámenes parciales 60 % Asistencia es obligatoria para aprobar el curso	Los exámenes constan de dos partes: una de resolución de problemas y la otra de descripción de conceptos. En ésta se le pide al alumno que dé el significado y la definición de conceptos propios de los temas de la asignatura. Dado que se está aprendiendo un lenguaje especializado, que depende de la materia que se imparta.
Examen 20% Tareas 50% Examen a casa 30%	El examen de clase constaba de 4 preguntas. El alumno tenía que responder tantas preguntas como le fuera posible. Una de las preguntas era escribir un ensayo de por lo menos dos cuartillas. En el examen que realizo para clase al final del curso pido que los alumnos escriban un ensayo sobre algún tema. Es una pregunta directa, pero me hace ver si el alumno es capaz de explicar con sus propias palabras algún tema físico, sin matemáticas (o pocas). Esta parte podía ser evadida por el estudiante pues tenía a elegir varias preguntas. Los estudiantes que mejor salen eligen esta opción porque resulta que es la pregunta más sencilla del examen.

5. ¿Cuáles fueron sus criterios para evaluar los trabajos, tareas o exámenes escritos de sus estudiantes?

3-4 y aplicación correcta de los conceptos	Nomenclatura empleada: 1 – Relevancia del contenido 2 – Uso apropiado de las fuentes 3 – Capacidad de reflexión 4 – Coherencia y cohesión 5 – Precisión gramatical 6 – Presentación y formato de acuerdo al género discursivo T – Todas las anteriores
1-2-3-4-5	
Exactitud matemática y argumentación clara	
3 y capacidad de análisis abstracto y desarrollo de ideas nuevas	
3 y entendimiento de los conceptos	
1-3-4-6	
1-2-3-4	
T	
1-3-4	
1-2-3-4-5 comprensión de conceptos	

Área 2: Ciencias Biológicas, Químicas y de la Salud
--

3. ¿Cómo evaluó a sus alumnos y qué peso total le otorgó a cada sección?
4. En caso de haber evaluado con alguna tarea/examen/trabajo escrito, describalo brevemente y mencione el número de cuartillas o palabras solicitadas en cada uno de ellos.

PRÁCTICAS DE EVALUACIÓN DOCENTE	ESCRITURA
Asistencia y participación en clase 20% Presentaciones de los temas 70 % Evaluación escrita 10%	Reseña sobre los nuevos temas aprendidos durante este curso. Una o dos páginas.
Presentación de proyecto de elaboración de una prueba 100%	Se solicita contenido no cantidad de páginas o palabras
Examen 10 % Asistencia y participación en clase 10 % Presentación oral en clase 50% Tareas 30 %	Siempre preparo un tarea (o examen) con 80 % de formato selección (seleccionar una respuesta correcta de varias posible respuestas) y 20 % de escribir la respuesta. En caso de pedir un escrito, no pido ninguna restricción de número de palabras. En general, pido a los alumnos que escriban de manera muy breve y concreta.
Participación 20% Exámenes 80%	Los exámenes están basados en el desarrollo de temas en formato libre
Avances en su proyecto de tesis en base a su cronograma	Redacción de la tesis y de artículos científicos.
Asistencia (25%) Presentación de avances de la tesis de doctorado frente a grupo (75%)	NO HAY EXAMEN ESCRITO
Asistencia 20% Participación 30% Examen parcial 25%, Presentación oral y redacción ensayo a elección libre 25%	Ensayo en el que desarrollen un tema del curso relacionado con su trabajo de tesis, máximo de 10 cuartillas, incluyendo bibliografía Exámenes parciales, con preguntas de opción múltiple y abiertas a desarrollar en máximo 4 renglones.
Asistencia 10% Participación 25% Exámenes parciales 40% Presentación 25%	Los exámenes que utilizamos fueron de opción múltiple y con imágenes.
Exámenes parciales 80%, participación y tareas 20%.	El examen parcial consiste de preguntas teóricas y prácticas de los temas revisados en clase. Generalmente de 10-12 preguntas para que tengan tiempo de resolverlo en clase. Las preguntas teóricas se contestan en un par de oraciones o a veces uso preguntas de opción múltiple. Las preguntas prácticas requieren que los alumnos resuelvan problemas.
Asistencia y participación 50%, Revisión y exposición de artículos 30% Tareas 20%	Tareas: lectura crítica de algunos artículos de un tema específico y realizar un trabajo escrito sobre el tema revisado. Longitud 5 cuartillas.
Trabajo Final 40% Examen 40% Participación 20%	El examen es único para todo el semestre, se entrega para contestar en casa a lo largo de una semana, pudiendo consultar cualquier material que dispongan y se promueve la participación y discusión grupal para resolver problemas que requieran la utilización de todos los conocimientos adquiridos.

5. ¿Cuáles fueron sus criterios para evaluar los trabajos, tareas o exámenes escritos de sus estudiantes?

1-2-3-4-6	<p>Nomenclatura empleada:</p> <p>1 – Relevancia del contenido</p> <p>2 – Uso apropiado de las fuentes</p> <p>3 – Capacidad de reflexión</p> <p>4 – Coherencia y cohesión</p> <p>5 – Precisión gramatical</p> <p>6 – Presentación y formato de acuerdo al género discursivo</p> <p>T – Todas las anteriores</p>
T	
1-4	
1-3-4	
1-2-4-6	
T	
Respuestas correctas	
Respuestas correctas	
1-2-4	

Área 3: Ciencias Sociales

3. ¿Cómo evaluó a sus alumnos y qué peso total le otorgó a cada sección?
4. En caso de haber evaluado con alguna tarea/examen/trabajo escrito, descríballo brevemente y mencione el número de cuartillas o palabras solicitadas en cada uno de ellos.

PRÁCTICAS DE EVALUACIÓN DOCENTE	ESCRITURA
60% trabajo final (avance de tesis) 40% participación y tareas	Avance de tesis
Para tener derecho a evaluación cumplir con el 80% de asistencia. 3 evaluaciones parciales: entregar las partes del proyecto terminadas (primera parte: problema, objetivos e hipótesis, calendario de trabajo; segunda parte: justificación, bibliografía mínima, conceptos y categorías clave, metodología general; tercera parte: marco teórico conceptual. La calificación final resulta del promedio de cada entrega.	En ninguno de mis cursos acostumbro evaluar con trabajo final. Prefiero mantener el ritmo de trabajo y atención en forma constante a lo largo del curso. Así que habitualmente hago evaluaciones parciales.
Asistencia y participación 15% Examen parcial 25% Casos y ejercicios 30% Proyecto de investigación 30%	Variaba mucho. Promedio 10 cuartillas
Asistencia 30% Participación en clase 30% Elaboración de trabajo con presentación de avances parciales y final oral 40%	El trabajo era de mínimo 50 pags. ya que era en equipo y cada alumno tenía un tema, las conclusiones se realizaban conjuntamente.
Trabajo final 50% Participación (asistencia, reportes y participación) 50%	15 cuartillas como máximo
Ensayo final, 50% Examen, 40% Exposiciones, 10%	Ensayo sobre algún tema del curso, desarrollado en 15 cuartillas con la biografía que se empleó en la clase.
Asistencia y participación 20% Presentación de un tema 40% Trabajo final 40%	Se pidió un trabajo final sobre uno de los temas del curso. Debía contener una crítica razonada de la bibliografía referida al tema central elegido. Extensión mínima de trabajo 20 páginas.
Resolución de casos prácticos 30% Participación en grupos de discusión 30% Participación en actividades extra-clase 10% Análisis de artículos de investigación 10% Participación en clase 10% Opinión del profesor 10%	Se resolvieron 10 casos prácticos (uno semanal), en dos cuartillas cada uno, con tres secciones obligatorias: 1) síntesis del caso, 2) respuesta a preguntas específicas y 3) opinión persona sobre el caso. El análisis sobre un artículo de investigación sobre administración de R.H. publicado en un journal internacional se reportó con extensión libre, tocando obligatoriamente 3 secciones: 1) síntesis del artículo 2) Explicación de la metodología usada por el (la) autor(a) 3) Opinión razonada acerca del artículo.

Asistencia y participación 30% Presentaciones orales 30% Ensayo final 40%	El ensayo final consta de un ensayo crítico sobre la lectura de un libro previamente asignado. La extensión es de un mínimo de 5 cuartillas y máximo de 15, escrito en letra times a 1.5 espacio.
Asistencia y participación 25% Ejercicios durante el curso 25% Trabajo final 50%	Se solicitó un ensayo final, de corte académico, entre 20 y 25 cuartillas a doble espacio con referencias bibliográficas.
40% de participación en el Seminario 30% en los reportes de lecturas y los análisis de contenido del material seleccionado para la materia 30% para un trabajo final.	Para un trabajo final de máximo quince cuartillas a doble espacio, desarrollar que aportará epistemológicamente el estudio de la cultura en la Antropología de las Empresas de cada estudiante del seminario. Se añadirá una cuartilla con una evaluación personal del trabajo en el seminario.

5. ¿Cuáles fueron sus criterios para evaluar los trabajos, tareas o exámenes escritos de sus estudiantes?

1-2-6 Metodología y datos originales empleados 2-4-6 y que cada parte del proyecto corresponda en el contenido (que el objetivo sea un objetivo, que el problema sea un problema, etc.). Dejé de evaluar gramática hace varios años porque hubieran reprobado muchos alumnos. Sin embargo marco los errores en tus textos.	<p>Nomenclatura empleada:</p> <p>1 – Relevancia del contenido 2 – Uso apropiado de las fuentes 3 – Capacidad de reflexión 4 – Coherencia y cohesión 5 – Precisión gramatical 6 – Presentación y formato de acuerdo al género discursivo T – Todas las anteriores</p>
T	
1-2-3-4	
1-2-3-4-5	
2-4-5	
1-2-3	
2-3-4-5-6	
1-2-3-4-6 y análisis de contenido y manejo del aparato crítico	
T	

Área 4: Humanidades y Artes

3. ¿Cómo evaluó a sus alumnos y qué peso total le otorgó a cada sección?
4. En caso de haber evaluado con alguna tarea/examen/trabajo escrito, descríballo brevemente y mencione el número de cuartillas o palabras solicitadas en cada uno de ellos.

PRÁCTICAS DE EVALUACIÓN DOCENTE	ESCRITURA
Participación y asistencia 20% Presentación oral de un tema 20% Trabajo escrito final 60%	El trabajo escrito fue el resultado de una investigación personal acerca de un tema, el cual fue presentado oralmente por el alumno en clase durante el semestre. 15 pp. a doble espacio, anotado y con bibliografía.
Enunciación del problema de investigación 10% Selección de fuentes documentales 20% Análisis de las fuentes seleccionadas 20% Ensayo final 50%	El ensayo final se va construyendo desde el inicio del semestre y los avances se discuten cada mes. El objetivo principal consiste en analizar una obra arquitectónica, incorporando los conceptos revisados en el aula.
Asistencia y participación 50% investigación de campo 25% y presentación de trabajo final 25%	En mi caso, se requiere la comprensión del idioma español al 100%
Texto 50%, Examen 50%	Se les pidió un texto de 15 cuartillas sobre alguno de los temas que se trataron durante el curso y se les hizo examen oral de análisis y crítica del periodo que se estudió.
Asistencia y participación 10% Presentación oral final 30 % Ensayo final 60 %	El ensayo final debía constar de un mínimo de diez páginas, más bibliografía.
Asistencia y participación 25% Ensayos (2) 25 % Análisis crítico de publicaciones 30% Asistencia a eventos académicos específicos 20%	En este curso se ofrece un rango variado de lecturas en historia, filosofía, sociología y crítica literaria. La metodología pedagógica en todos los cursos es la siguiente: mesa de debate crítico sobre lecturas seleccionadas. El alumno debe a) desarrollar una lectura y opinión críticas sobre el material solicitado, b) presentar argumentos en mesa de debate (horas clase) y c) sintetizar en un pequeño texto las ideas y argumentos individuales y grupales (10 cuartillas)
Trabajos y cuestionarios 40% Presentación de una secuencia didáctica 40% Examen parcial 20%	Los trabajos tuvieron una extensión de cuatro cuartillas. Los cuestionarios eran de cinco preguntas. La presentación de la secuencia didáctica constó, en la mayoría de los casos de 20 cuartillas incluida la bibliografía.
La presentación crítica de uno de los escritores estudiados en clase (20%), presentación al final del curso de un ensayo sobre un cuentista (60%) y la participación en clase (20%)	Se examinó la pertinencia del enfoque crítico aplicado, la pertinencia del tema elegido y la profundidad del examen hecho.
Asistencia (5%), lecturas y participación en clase (10%), presentación de un tema al grupo (15%), un primer trabajo escrito (20%), y un ensayo final (50%).	El ensayo final fue de 10 a 14 cuartillas. Cada estudiante definió su propio tema de investigación que trabajó a lo largo del semestre con base en algunas de las discusiones en clase y los textos que estudiamos.

Participación en clase, trabajos escritos y presentaciones orales	Trabajos escritos semanales de una a dos cuartillas. Trabajo final 10 a 15 cuartillas
100% presentación de los resultados de una investigación de tipo cualitativo.	Se pidió al alumno que en 25 cuartillas presentara los resultados de su trabajo.
80% entrega de ensayo 20% exposición oral y participación.	En cada sesión los alumnos debían entregar una cuartilla de reseña sobre el tema a revisar (no tenía una calificación precisa, pero era evaluada la escritura y pre-requisito para estar y aprobar el seminario).
Asistencia 33% participación 33% ensayo final 33%	En el ensayo final se pidió que se abordaran los conceptos teóricos estudiados en clase y que se desarrollaran en relación con un texto literario leído de manera independiente por los alumnos en un espacio de entre 15 y 20 cuartillas. Una alumna extranjera recibió las instrucciones con anterioridad para poder pulir su español.
Asistencia y participación en clase: 25% Exposición individual: 25% Examen oral final: 25% Trabajo final: 25%	El trabajo consistió en un capítulo de la tesis de maestría o doctorado que cada alumno esté realizando. A veces, por hallarse en los comienzos de su investigación, dicho capítulo es en realidad el proyecto explícito del trabajo que habrán de realizar.
solo se pone acreditado o no, cumplimiento de actividades 50%, entrega de avances de tesis 50%	Medio capítulo de tesis y esquema de trabajo para el semestre siguiente.
Exposiciones 60%, Asistencia 20% y Lecturas/participación 20%	
Asistencia y participación 40% Trabajo final 60%	Trabajo escrito con una extensión de quince a veinte páginas.
Asistencia y participación 20% Presentación de un tema libre 40% Trabajo final 40%	En cuanto al trabajo solicitado, no hubo una referencia cuantitativa en cuanto a la extensión. Más bien solicité una discusión relacionada con el tema de su tesis de grado, que emplearan los códigos como fuente indígena escrita. Es claro que hay temas que tienen una extensión importante, otros, apenas y se refieren a una o dos imágenes, que son los menos frecuentes.
Asistencia y participación 60% Trabajo final 40%	El trabajo final consistió en hacer un ejercicio de grabación, procesamiento y análisis de datos lingüísticos en distintas lenguas distintas al español.
Ensayo final 90% Asistencia y participación 10%	Fue un ensayo final, de 12 a 20 cuartillas.
Entrega y exposición 20% Contenido 20% Cohesión y coherencia 20% Asistencia 20% Participación 20%	El semestre se evaluó con la entrega de tres trabajos escritos con base en los proyectos de investigación para su tesis de maestría. El primer trabajo consistió en la creación de un párrafo con el planteamiento del problema y la hipótesis de trabajo. El segundo fue la escritura de una cuartilla cuyo contenido se refiriera a una parte del proyecto de investigación; se sugirió que escogiesen aquella parte de la que ya podrían presentar resultados. El tercer y último trabajo consistió en la reunión adecuada de las dos partes antes escritas, con la posibilidad de extenderse a cinco cuartillas.
Proyecto de investigación 80%	Detallar el proyecto de investigación, no había límite de palabras, sólo máximo 15 cuartillas.

Lectura de artículos y libros - incorporado al marco teórico- (30%), diseño de un instrumento de captación de información para ser aplicado a la población meta de la investigación (30%),	Debido a que el Seminario de Investigación promueve que los estudiantes redacten su investigación, las lecturas sugeridas o investigadas individualmente se tienen que incorporar en el capítulo del Marco Teórico de su investigación. En la elaboración del instrumento se pide lo diseñen con las preguntas pertinentes al tema y que reflejen la metodología a seguir.
Exposición de un tema en clase, con presentación en Power Point: 25% Ensayo final: 40%	Para el PPT la parte escrita es obviamente poca. El ensayo final es de entre 5 y 10 páginas. Tienen que explicar qué tema les gustaría trabajar a partir de la teoría de las representaciones sociales: por qué les interesa ese tema, qué aportaría al área, con qué metodología (instrumentos, sujetos, tiempo), por qué es un tema de representaciones sociales, reflexiones personales finales. Lo hacen en casa
Reportes gráficos y escritos (50%)	El reporte escrito de 20 cuartillas aproximadamente 400 palabras pues incluía material gráfico se elaboró durante 7 semanas. Fue entregado y presentado oralmente.

5. ¿Cuáles fueron sus criterios para evaluar los trabajos, tareas o exámenes escritos de sus estudiantes?

T	<p>Nomenclatura empleada:</p> <p>1 – Relevancia del contenido 2 – Uso apropiado de las fuentes 3 – Capacidad de reflexión 4 – Coherencia y cohesión 5 – Precisión gramatical 6 – Presentación y formato de acuerdo al género discursivo T – Todas las anteriores</p>
1-2-3-4-5 y el nivel de análisis de la obra	
1-2-4-6	
T	
T	
1-3-4 y constancia de investigación activa	
1-3-4 y creatividad intelectual	
1-2-3-4	
1-2-3-4	
1-2-3-4-5	
T	
T	
1-2-3-4-5	
T	
1-2-3-4	
2-3-4	
1-2-3	
1-2-3	
1-2-3-4-6	
T	
T y correspondencia con el tema de investigación	

Anexo 1.5

Comentarios adicionales de los docentes de posgrado

Neurobiología

El uso correcto del idioma es muy importante para los estudiantes de posgrado. Sin embargo, el nivel que observamos en general es bajo y muy heterogéneo porque vienen de diferentes instituciones y carreras profesionales.

Estudios Mesoamericanos

En general, se espera que un estudiante que ingresa a la maestría cuente con las herramientas suficientes y adecuadas de lecto-escritura, pues ya ha cursado anteriormente la licenciatura y, en la mayor parte de los casos, ha defendido una tesis de grado.

Astrofísica

Una gran parte de los estudiantes hispanoparlantes presentan enormes deficiencias en expresión escrita en español. Aunque la mayor parte de la comunicación escrita en la astrofísica se realiza en inglés, me parece que las deficiencias en la capacidad de redacción en español provocan que la redacción en inglés se dificulte aún más. Creo que las dificultades mayores no provienen de la falta de conocimiento del español, sino de la falta de conocimiento del lenguaje en general y de la falta de capacidad para sintetizar información, ordenar ideas y expresarse precisamente. Por otro lado, también es deseable que los egresados del posgrado puedan redactar textos (proyectos, reportes, etc.) en español y eso resulta sumamente difícil para quienes no saben redactar (independientemente de su lengua materna). Creo que si se ofrecen cursos de redacción para estudiantes de posgrado estos también deberían ofrecerse a estudiantes hispanohablantes.

Ciencia Bioquímica

Ya había comentado anteriormente, encuentro frecuentemente una respuesta con muchas decolaciones no necesario para una pregunta de tarea o examen. Pienso que muchos alumnos tienen idea de que el número de palabra mayor sea lo mejor. Pienso que esto es incorrecto en particularmente en campo de ciencia natural.

Ingeniería de sistemas

Creo que para un extranjero es difícil manejar correctamente los aspectos del español que son diferentes que en su propio idioma (por ejemplo, género, subjuntivo, diferencia entre ser y estar...). Sin embargo, si la estructura de los escritos está bien en su idioma, estará generalmente bien en español. Una falta de estructura en lo que escribimos es según yo nada más que la falta de estructura en las ideas que tenemos....

Posgrado en Ciencia e Ingeniería de la Computación

Gracias por solicitar la opinión de quienes damos clase en el Posgrado. ¿Han pensado en algo similar para evaluar a candidatos a profesores cuya lengua materna no es el español?

Ciencias Biológicas

Las preguntas me parecen muy acertadas y las respuestas solo indican un deber ser, desgraciadamente a este nivel los alumnos en su mayoría carecen de un buen manejo del lenguaje y su vocabulario es pobre. La redacción en general no es muy buena y estoy clara de la necesidad de que ellos cursen por un pequeño curso de gramática que les permita mejorar estos aspectos. Muchas gracias.

Ingeniería

Además de los trabajos escritos que se solicitan en clase, los alumnos de maestría deben realizar artículos técnicos para congresos y desarrollar una tesis de grado.

Ciencia e Ingeniería de Materiales

Es importante que los alumnos lleven cursos de redacción, composición, ortografía y puntuación en cada uno de los niveles de estudio: licenciatura, especialidad, maestría y doctorado.

Posgrado en Historia del Arte, de la Facultad de Filosofía y Letras

Me resulta de mucho interés este tipo de encuestas.

MADEMS

Los alumnos extranjeros que he tenido son de nivel licenciatura y cursan un semestre o un año por movilidad estudiantil.

Antropología

Curiosamente, los alumnos extranjeros no presentaron problemas de escritura

Astrofísica

Es muy poco el material que se escribe en español para investigación científica. Los artículos y ediciones casi todos tienen que ser en lengua inglesa. Casi todo lo que se hace en español es para la difusión local de la ciencia que realizamos. En términos generales no es necesario que un alumno de maestría (y menos de doctorado) en ciencias sepa mucho español (aunque es deseable). Al ritmo que llevamos, incluso las tesis ahora pueden ser un 90% en inglés (solamente se traduce al español la introducción, título y conclusiones). Así es que no dudo que en un futuro no muy lejano incluso las clases en los posgrados de astrofísica y física sean en inglés.

Lingüística aplicada

La mayoría de los conceptos son muy importantes si se quiere que el texto sea del nivel académico adecuado a una maestría cualquiera.

Lingüística aplicada

En el Seminario de Investigación la evaluación es más cualitativa que cuantitativa. Tal vez no se vea el producto (investigación) terminado hasta que se termine la misma, sin embargo, sí se aprecian los cambios y reflexiones que se van incorporando.

Arquitectura

En el área de arquitectura y en particular en materiales como tecnología y diseño generalmente, aunque no siempre, se da más peso a la creatividad la cual es evaluada a través de material gráfico. Sin embargo, creo que es muy importante el que estas ideas se expresen también a través de textos bien redactados.

ANEXOS ESTUDIO 2

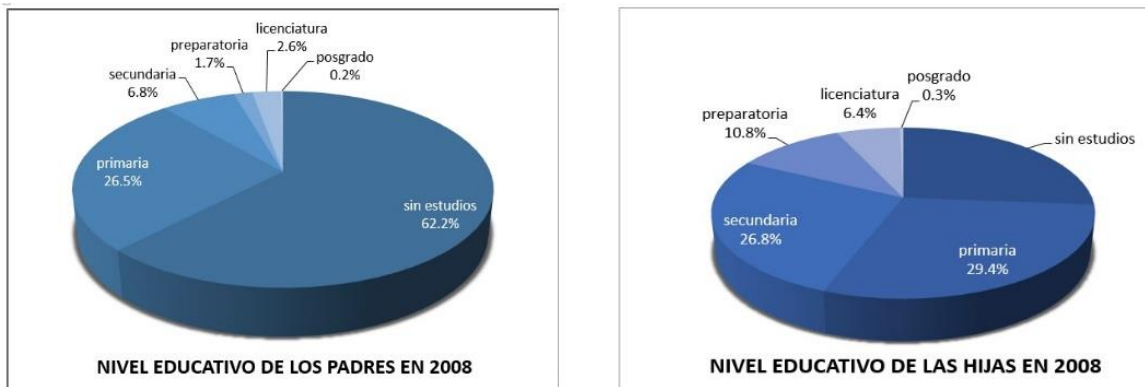
Anexo 2.1

Prueba escrita del EXELEAA (modelo 1)

Prueba 3. Expresión escrita

La prueba de Expresión Escrita consta de DOS tareas. Tienes 70 minutos para realizarlas.

Tarea 1. A continuación se presentan dos gráficas sobre el logro educativo que han alcanzado en la última década las hijas con respecto a sus padres en un país de América Latina. En el cuadro de texto que aparece debajo de las gráficas, descríbelas y contrástalas utilizando mínimo 150 palabras. Toma 25 minutos para esta tarea.



Tarea 2

a) Lee la siguiente información: “Si bien la educación por sí misma amplía los horizontes de los individuos y constituye un beneficio para éstos, la situación educativa de hombres y mujeres tiene diferentes beneficios para unos y otros y para sus hijos e hijas” (Milosavljevic, 2007: 96).

b) En el cuadro de texto que aparece en la parte inferior, escribe un ensayo sobre el tema de la educación de las mujeres. Puedes retomar información de la Tarea 1.

- Menciona la importancia del tema.
- Establece una postura al respecto.
- Establece los pros y los contras de la temática.
- Menciona lo que ocurre en tu país.
- Concluye.

Escribe mínimo 250 palabras. Tienes 45 minutos para realizar esta tarea.

Anexo 2.2

Rúbrica para evaluar las 2 tareas que conforman la prueba de expresión escrita del EXELEAA (primera versión)

Competencia sociolingüística	/ 3
Competencia estratégica	/ 3
Competencia lingüística	/15
Competencia pragmático-discursiva	/21
Valoración global	/ 3
TOTAL	/45 = 20%

Competencia sociolingüística (3 puntos)

ADECUACIÓN

	Puntos	T1	T2	Total
Sin ninguna falla en el uso de las convenciones formales académicas : título, registro formal, tratamiento impersonal, objetividad (evitar anécdotas y frases como <i>yo creo que...</i>)	3			/6
Hasta 2 de las siguientes fallas de adecuación: ausencia de título (o utilización del mismo como introducción al primer párrafo), registro informal (uso de coloquialismos, barbarismos, palabras truncadas, etc.), inconsistencia en el tratamiento (cambio de persona gramatical), inserción de anécdotas personales y frases como <i>yo creo que...</i>	2			
Inadecuado (más de 2 fallas de adecuación).	1			

Competencia estratégica (3 puntos)

	Puntos	T1	T2	Total
Planeación y monitoreo (o un texto sin evidencia de estrategias, pero muy bien escrito)	3			/6
Planeación (esquemas, diagramas, lluvia de ideas).	2			
Monitoreo (reparación, reajuste de la tarea).	1			
Sin evidencia de estrategias en un texto fallido.	0			

Competencia lingüística (15 puntos)

RANGO Y CONTROL LÉXICO

	Puntos	T1	T2	Total
Vocabulario variado (incluye sinónimos), preciso y con la inserción de algunos términos especializados en relación con el tema de la tarea. Menos de 3 errores.	3			/6
Vocabulario adecuado para el contexto de la tarea (puede presentar algunas repeticiones, redundancias o expresiones imprecisas, como palabras fáciles). 3 errores.	2			
Vocabulario pobre y repetitivo. Más de 3 errores.	1			

PRECISIÓN GRAMATICAL

	Puntos	T1	T2	Total
Sin errores en el uso de verbos, adverbios, pronombres, preposiciones, conjunciones, gerundios y locuciones.	3			/6
Hasta 3 errores en el uso de verbos, adverbios, pronombres, preposiciones, conjunciones, gerundios y locuciones.	2			
Más de 3 errores en el uso de verbos, adverbios, pronombres, preposiciones, conjunciones, gerundios y locuciones.	1			

COMPLEJIDAD SINTÁCTICA

	Puntos	T1	T2	Total
Uso de 5 o más nexos subordinantes distintos, sin errores (p. ej. mal empleo de nexo o problemas de correlación entre ellos). Se incluyen otros nexos además de <i>que, ya que, porque</i> y <i>aunque</i> .	3			/6
Uso de 4 nexos subordinantes distintos, sin error (o 5 nexos subordinantes distintos hasta con un error). Uso de los nexos más característicos (<i>que, ya que, porque, aunque</i>).	2			
Uso de 3 o menos nexos subordinantes distintos sin errores.	1			
Uso de 3 o menos nexos subordinantes distintos con errores.	0			

ORTOGRAFÍA

	Puntos	T1	T2	Total
Sin errores ortográficos.	3			/6
Hasta 3 errores ortográficos (en palabras distintas).	2			
Más de 3 errores ortográficos (en palabras distintas).	1			

PUNTUACIÓN

	Puntos	T1	T2	Total
Sin errores de puntuación e inclusión de signos como <i>¡! ¿? ; : () — ... “”</i>	3			/6
Uso correcto del punto (.) y de la coma (,), pero con errores en el uso de signos como <i>¡! ¿? ; : () — ... “”</i>	2			
Uso correcto del punto (.) y de la coma (,)	1			
Errores de puntuación incluso en el uso del punto y coma.	0			

Competencia pragmático-discursiva (21 puntos)

DESARROLLO DE IDEAS Y PRECISIÓN PROPOSICIONAL

	Tarea 1	
	Sí (1 pto.)	No (0 pto.)
¿Presenta la descripción de los elementos que componen la gráfica ?		
¿Describe la información relevante ?		
¿Describe la información del resto de la tabla ?		
¿Hay una conclusión que resume y establece relación entre los datos?		
TOTAL		/4

Nota: No se valora la interpretación de los datos.

	Tarea 2	
	Sí (1 pto.)	No (0 ptos.)
¿Hay un título relacionado con el tema a tratar?		
¿Hay una introducción que presente el marco temático ?		
¿Enuncia claramente la tesis?		
¿Presenta al menos dos argumentos?		
¿Los argumentos son objetivos y pertinentes?		
¿Reconoce al menos un contraargumento?		
¿Hay una conclusión que sintetiza el planteamiento o reelabora la tesis? (También se obtiene 0 si el texto tiene cierre, pero no conclusión)		
¿Incluye alguna de las siguientes operaciones? <ul style="list-style-type: none"> • Secuencia progresiva. • Ejemplificación. • Definición. • Clasificación. • Comparación o contraste. 		
¿Incluye alguna de las siguientes operaciones? <ul style="list-style-type: none"> • Analogía interesante. • Cita de autoridades. • Cita de datos objetivos. • Pregunta o figura retórica (metáfora, metonimia, repetición enfática, etc.) • Ironía. • Golpe de efecto en las conclusiones. 		
TOTAL		/11

COHERENCIA

	Puntos	T1	T2	Total
Hilo conductor claro entre los distintos párrafos y entre las ideas. Uso variado y apropiado de marcadores discursivos (<i>por un lado, por otro lado; asimismo; en primer lugar, en segundo lugar; en otros términos; etcétera</i>).	3			/6
Ideas y párrafos unidos de manera elemental . Uso limitado de marcadores discursivos (<i>además, también, por eso, es decir, o sea, esto es, etcétera</i>). Hasta 3 errores.	2			
Ideas y párrafos incoherentes que incluso llegan a romper el principio de la no contradicción. Uso incorrecto de los marcadores discursivos más comunes o ausencia de ellos. Más de 3 errores.	0			

COHESIÓN

	Puntos	T1	T2	Total
Uso apropiado de recursos de referencia interna (conjunciones, concordancias, deícticos, elipsis, que evitan repeticiones innecesarias).	3			/6
Uso elemental de recursos de referencia interna (conjunciones, concordancias, deícticos, elipsis), que provoca la aparición de hasta tres repeticiones innecesarias.	2			
Ausencia de recursos de referencia interna (conjunciones, concordancias, deícticos, elipsis) o más de tres repeticiones innecesarias.	1			

Anexo 2.3

Matriz de evaluaciones con la primera versión de la rúbrica

Examinado	E-1 ¹	E-2	E-3	E-4	E-5
1	T1 y T2 ²	T1 y T2			
2	T1 y T2			T1 y T2	
3	T1 y T2		T1 y T2		
4	T1 y T2			T1 y T2	
5	T1 y T2	T1 y T2			
6	T1 y T2	T1 y T2			
7	T1 y T2	T1 y T2	T1 y T2		
8		T1 y T2	T1 y T2		
9		T1 y T2		T1 y T2	
10		T1 y T2	T1 y T2		
11		T1 y T2		T1 y T2	
12		T1 y T2	T1 y T2		
13		T1 y T2		T1 y T2	
14		T1 y T2	T1 y T2		
15		T1 y T2	T1 y T2		
16			T1 y T2	T1 y T2	
17			T1 y T2	T1 y T2	
18			T1 y T2		T1 y T2
19			T1 y T2		T1 y T2
20	T1 y T2	T1 y T2			
21	T1 y T2	T1 y T2			
22	T1 y T2	T1 y T2			
23	T1 y T2	T1 y T2			
24	T1 y T2	T1 y T2			
25	T1 y T2	T1 y T2			
26	T1 y T2	T1 y T2			
27	T1 y T2	T1 y T2			
28	T1 y T2	T1 y T2			
29	T1 y T2	T1 y T2			
30	T1 y T2	T1 y T2			
31	T1 y T2	T1 y T2			
32	T1 y T2	T1 y T2			
33	T1 y T2	T1 y T2			
34	T1 y T2	T1 y T2			
35	T1 y T2	T1 y T2			
36	T1 y T2	T1 y T2			
37	T1 y T2	T1 y T2			

¹ E= evaluador² T1= Tarea 1, T2= Tarea 2, T3= Tarea 3 y T4= Tarea 4

Anexo 2.4

Estudiantes universitarios que participaron en la prueba escrita del EXELEAA

Examinado	Universidad de procedencia	Grado	Nacionalidad
1	Universidad Nacional Autónoma de México	Posgrado	Italiana
2			Checa
3			Italiana
4			Canadiense
5			italiana
6			Belga
7			italiana
8			Checa
9			estadounidense
10			Italiana
11			Japonesa
12			Japonesa
14			Estadounidense
15			Francesa
16			Belga
17			estadounidense
18			francesa
19			estadounidense
20			Austriaca
21			inglesa
22			Francesa
23		Brasileña	
24		Alemana	
25		Rumana	
26		Francesa	
27		Alemana	
28		Noruega	
29		canadiense	
30		Francesa	
31		Alemana	
32		Alemana	
33		Austriaca	
34		Francesa	
35		Brasileña	
36		Austriaca	
37		Alemana	
38		Francesa	
39		alemana	
40		Universidad de Guadalajara	Grado
41	s/d		
42	s/d		
43	s/d		
44	s/d		
45	francesa		
46	Estadounidense		
47	francesa		
48	francesa		
49	japonesa		
50	brasileña		

51			alemana
52			francesa
53			francesa
54			francesa
55			francesa
56			taiwanesa
57			francesa
58			francesa
59			irlandesa
60			taiwanesa
61			japonesa
62			japonesa
63			japonesa
64			alemana
65			israeli
66			Polaca
67			belga
68			alemana
69			Alemana
70			Checa
71			alemana
72			trinidad
73			alemana
74			italiana
75			alemana
76			china
77			surcoreana
78			italiana
79			estadounidense
80			china
81			china
82			alemana
83			estadounidense
84			italiana
85			japonesa
86			inglesa
87			canadiense
98			alemana
99			japonesa
88			estadounidense
89			estadounidense
90			estadounidense
91			estadounidense
92			estadounidense
100			estadounidense
13			estadounidense
93			s/d
94			s/d
95			s/d
96			s/d
97			s/d
	Universidad Nacional Autónoma de México	Posgrado	
	Universidad Panamericana	Grado	
	Instituto Tecnológico de Estudios Superiores Monterrey CCM	Grado	

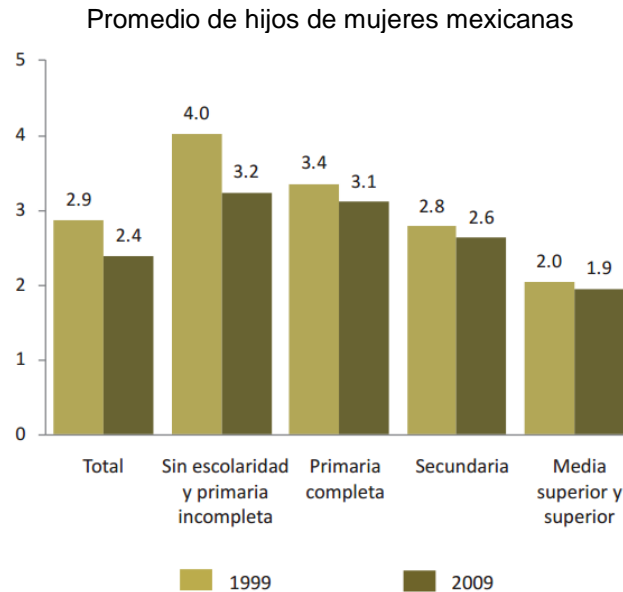
Anexo 2.5

Prueba escrita del EXELEAA (modelo 2)

Prueba 3. Expresión escrita

La prueba de Expresión Escrita consta de DOS tareas. Tienes 70 minutos para realizarlas.

Tarea 1. A continuación se te presenta una gráfica de barras. En el cuadro de texto que aparece debajo de las gráficas, descríbelas y contrástalas utilizando aproximadamente 200 palabras. Tienes aproximadamente 25 minutos para esta tarea.



Fuente: INEGI. XII Censo General de Población y Vivienda 2000. Base de datos; Censo de Población y Vivienda 2010. Base de datos de la muestra.

Tarea 2

En algunos países existen solamente sistemas de salud pública, mientras que en otros existen sistemas públicos y privados. Con base en dicha temática, escribe un ensayo argumentativo con base en lo siguiente:

- Menciona la importancia del tema.
- Establece una postura al respecto.
- Establece los pros y los contras de los sistemas públicos y los privados.
- Menciona cómo es el sistema de salud pública en tu país.
- Concluye.

Escribe mínimo 300 palabras. Tienes aproximadamente 45 minutos para realizar esta tarea.

Anexo 2.6

Rúbrica para evaluar las 2 tareas que conforman la prueba de expresión escrita del EXELEAA (segunda versión)

Competencia sociolingüística	/6
Competencia pragmático-discursiva	/18
Competencia lingüística	/18
TOTAL	/42

Competencia sociolingüística

Adecuación: enfoque en el tema y uso de las convenciones formales.		
MARCA	PTOS	DESCRIPCIÓN
EXCEPCIONAL	3	El discurso logra plenamente su propósito comunicativo . Sin fallas en el uso de las convenciones formales académicas : registro formal, tratamiento impersonal, objetividad (evitar anécdotas y frases como <i>yo creo que...</i>). Se adecua al número de palabras solicitado (en un rango de 10% mayor o menor).
HÁBIL	2	El discurso logra su propósito comunicativo , pero presenta alguna falla en cuanto a: Inserción de anécdotas personales y frases como <i>yo creo que...</i> ; cita textual de las palabras utilizadas en la consigna. No se ajusta al número de palabras solicitado (en un rango de 20% mayor o menor).
EN DESARROLLO	1	El discurso NO logra su propósito comunicativo . Varias fallas de adecuación: Inserción de anécdotas personales y frases como <i>yo creo que...</i> ; cita textual de las palabras utilizadas en la consigna. No se ajusta al número de palabras (en un rango menor a 50%)
	0	El discurso NO logra su propósito comunicativo . Texto <i>fuera de las convenciones formales</i> : registro informal (uso de coloquialismos, barbarismos, palabras truncadas, etc.), inconsistencia en el tratamiento (cambio de persona gramatical). Texto demasiado corto para ser evaluado.

Competencia pragmático-discursiva

Desarrollo de ideas (tarea 1). Esquemas que caracterizan los tipos de texto y que conducen su interpretación y producción (superestructura)		
MARCA	PTOS	DESCRIPCIÓN
EXCEPCIONAL	3	La redacción se centra en el tema propuesto. Se desarrollan completamente todos los núcleos temáticos . Describe los componentes de la(s) gráfica(s) (sin retomar literalmente las palabras de la consigna). Contrasta de manera general la información relevante. Sintetiza la información del resto de la tabla . Presenta una conclusión que resume y establece relación entre los datos.
HÁBIL	2	Se desarrollan en su mayoría los núcleos temáticos . El texto carece de alguno de los siguientes elementos: Descripción de los componentes de la tabla. Contraste de información relevante. Síntesis del resto de la tabla. Conclusión.

EN DESARROLLO	1	Se desarrollan parcialmente algunos de los núcleos temáticos . El texto carece de dos de siguientes elementos: Descripción de los componentes de la tabla. Contraste de información relevante. Síntesis del resto de la tabla. Conclusión. Presenta alguna de las siguientes fallas: Falta de abstracción del significado principal por descripción demasiado atomizada de los elementos. Inclusión de información que no viene en las gráficas. Interpretación de los datos.
	0	Se desarrolla aisladamente sólo alguno de los núcleos temáticos o aborda una temática diferente a la solicitada. Texto fallido por la acumulación de los siguientes elementos: Falta de abstracción del significado principal por descripción demasiado atomizada de los elementos. Inclusión de información que no viene en las gráficas. Interpretación de los datos.

Desarrollo de ideas (tarea 2). Esquemas que caracterizan los tipos de texto y que conducen su interpretación y producción (superestructura)		
MARCA	PTOS	DESCRIPCIÓN
EXCEPCIONAL	3	La redacción se centra en el tema propuesto. Se desarrollan completamente todos los núcleos temáticos . Hay una introducción que presente el marco temático. Enuncia claramente la tesis . Presenta al menos dos argumentos fundamentados. Reconoce al menos un contrargumento (o menciona que no los hay, según su punto de vista). Hay una conclusión que sintetiza el planteamiento o reelabora la tesis. Incluye alguna operación retórica -Secuencia progresiva -Ejemplificación -Definición -Comparación o contraste -Cita de autoridades -Pregunta retórica, etc.
HÁBIL	2	Se desarrollan en su mayoría los núcleos temáticos . El texto carece de alguno de los siguientes elementos: -Introducción -Tesis -Un argumento -Contrargumento -Conclusión -Operaciones
EN DESARROLLO	1	Se desarrollan parcialmente algunos de los núcleos temáticos . El texto carece de dos de los siguientes elementos: -Introducción -Tesis -Un argumento -Contrargumento -Conclusión
	0	Se desarrolla aisladamente sólo alguno de los núcleos temáticos o aborda una temática diferente a la solicitada. El texto carece de estructura argumentativa.

Cohesión. Relación entre oraciones, especialmente de referencia textual. Relaciones dentro del párrafo. Incluye puntuación (microestructura).		
MARCA	PTOS	DESCRIPCIÓN
EXCEPCIONAL	3	Uso experto de recursos de referencia interna (concordancias, deícticos, elipsis). Incluye léxico (como las nominalizaciones o los sinónimos) que retoman información y evitan repeticiones innecesarias en un párrafo. Uso variado y apropiado de los nexos sintácticos propios de la escritura académica (principalmente comparativo, causativo y/o concesivo). Uso preciso de los signos de puntuación (incluyendo signos como ¡! ¿? ; : () — ... “”).
HÁBIL	2	Uso apropiado de recursos de referencia interna (concordancias, deícticos, elipsis). Uso de los nexos sintácticos más comunes (<i>que, el cual, quien, ya que, porque, aunque quien/es, pero, sin embargo</i>). Correcto uso de los signos de puntuación.
EN DESARROLLO	1	Algunas fallas por la falta de recursos de referencia interna. Repeticiones. Uso de oraciones simples, coordinadas o yuxtapuestas y de algunas subordinadas con el nexo <i>que</i> . Predominio del punto y seguido para unir ideas.
	0	Múltiples faltas de concordancia y repeticiones causadas por la ausencia de recursos de referencia interna. Predominio del punto y seguido, con fallas en el uso de otros signos de puntuación.

Coherencia: se relaciona con las relaciones semánticas globales del contenido del texto que reflejan el sentido general del discurso. Uso de marcadores tanto dentro del párrafo, como de uno a otro (macroestructura)		
MARCA	PTOS	DESCRIPCIÓN
EXCEPCIONAL	3	Hilo conductor claro entre las ideas y entre los distintos párrafos. Uso variado y apropiado de marcadores discursivos propios de la lengua académica (<i>por un lado, por otro lado; asimismo; en primer lugar, en segundo lugar; en otros términos; mientras; en contraste, por otro lado, etcétera</i>).
HÁBIL	2	Hilo conductor entre las ideas y entre los párrafos. Uso de los marcadores más comunes (<i>además, también, es decir, o sea, esto es, etcétera</i>), principalmente dentro de un párrafo.
EN DESARROLLO	1	Estructuración en párrafos, pero sin uso de marcadores discursivos.
	0	Falta de estructuración en párrafos. Ausencia de marcadores.

Competencia lingüística

Rango y control léxico (sustantivos, adjetivos y verbos).		
MARCA	PTOS	DESCRIPCIÓN
EXCEPCIONAL	3	Vocabulario variado, preciso y con la inserción de algunos términos especializados en relación con el tema de la tarea.
HÁBIL	2	Vocabulario académico y adecuado para el contexto de la tarea (puede presentar algunas redundancias o expresiones calcadas de la lengua materna).
EN DESARROLLO	1	Vocabulario propio de los ámbitos público y privado . Repetición de palabras. Uso de palabras fáciles.
	0	El vocabulario es demasiado pobre para comprender las ideas o presenta excesivos calcos léxicos de la lengua materna.

Precisión gramatical. Uso de las categorías gramaticales, modos y tiempos verbales (no incluye concordancia).		
MARCA	PTOS	DESCRIPCIÓN
EXCEPCIONAL	3	Uso preciso de categorías gramaticales. Prácticamente sin errores.
HÁBIL	2	Uso adecuado de las categorías gramaticales, con algunas imprecisiones .
EN DESARROLLO	1	Errores en el uso de las categorías gramaticales que evidencian la estructura sintáctica de otra lengua .
	0	Múltiples errores en el uso de las categorías gramaticales que impiden la comunicación del mensaje .

Ortografía		
MARCA	PTOS	DESCRIPCIÓN
EXCEPCIONAL	3	Prácticamente sin errores ortográficos.
HÁBIL	2	Uso correcto de la ortografía, aunque con algunos errores (incluye los producidos por la interferencia de la lengua materna).
EN DESARROLLO	1	Constantes errores ortográficos.
	0	Errores ortográficos que impiden la comprensión de la lectura.

Anexo 2.7

Matriz de evaluaciones con la segunda versión de la rúbrica

Examinado	E1 ³	E2	E3	E4	E5	E6
1		T1 y T2	T1	T2	T1	T2
2		T1 y T2	T1	T2	T2	T1
3		T1 y T2	T1	T2	T1	T2
4		T1 y T2	T1	T2	T2	T1
5		T1 y T2	T1	T2	T1	T2
6		T1 y T2	T1	T2	T2	T1
7		T1 y T2	T1	T2	T1	T2
8		T1 y T2	T1	T2	T2	T1
9		T1 y T2	T1 y T2	T1 y T2	T1 y T2	T1 y T2
10		T1 y T2	T1 y T2	T1 y T2	T1 y T2	T1 y T2
11		T1 y T2	T1	T2	T1	T2
12		T1 y T2	T1 y T2	T1 y T2	T1 y T2	T1 y T2
13		T1 y T2	T1	T2	T1	T2
14	T1 y T2 ⁴	T1 y T2	T1 y T2	T1 y T2	T1 y T2	T1 y T2
15		T1 y T2	T1	T2	T2	T1
16		T1 y T2	T1	T2	T1	T2
17		T1 y T2	T1	T2	T2	T1
18		T1 y T2	T1	T2	T1	T2
19		T1 y T2	T1	T2	T2	T1
20		T1 y T2	T1	T2	T1	T2
21		T1 y T2	T1	T2	T2	T1
22	T1 y T2	T1 y T2	T1	T2	T1	T2
23	T1 y T2	T1 y T2	T2	T1	T2	T1
24	T1 y T2	T1 y T2	T2	T1	T1	T2
25	T1 y T2	T1 y T2	T2	T1	T2	T1
26	T1 y T2	T1 y T2	T2	T1	T1	T2
27	T1 y T2	T1 y T2	T2	T1	T2	T1
28	T1 y T2	T1 y T2	T2	T1	T1	T2
29	T1 y T2	T1 y T2	T2	T1	T2	T1
30	T1 y T2	T1 y T2	T2	T1	T1	T2
31	T1 y T2	T1 y T2	T2	T1	T2	T1
32	T1 y T2	T1 y T2	T2	T1	T1	T2
33	T1 y T2	T1 y T2	T2	T1	T2	T1
34	T1 y T2	T1 y T2	T2	T1	T1	T2
35	T1 y T2	T1 y T2	T2	T1	T2	T1
36	T1 y T2	T1 y T2	T2	T1	T1	T2
37	T1 y T2	T1 y T2	T2	T1	T2	T1
38	T1 y T2	T1 y T2	T2	T1	T1	T2
39	T1 y T2	T1 y T2	T2	T1	T2	T1
40	T2	T2	T2	T2	T2	T2
41	T1 y T2	T1 y T2	T1	T2	T2	T2
42	T1 y T2	T1 y T2	T1	T2	T2	T2
43	T2	T2	T2	T2	T2	T2
44	T2	T2	T2	T2	T2	T2
45	T2	T2	T2	T2	T2	T2
46	T2	T2	T2	T2	T2	T2
47	T2	T2	T2	T2	T2	T2
48	T2	T2	T2	T2	T2	T2
49	T1 y T2	T1 y T2	T1	T2	T2	T2

³ E= evaluador⁴ T1= Tarea 1, T2= Tarea 2, T3= Tarea 3 y T4= Tarea 4

Candidato	E1	E2	E3	E4	E5	E6
50	T1 y T2	T1 y T2	T1	T2	T2	T2
51	T2	T2	T2	T2	T2	T2
52	T2	T2	T2	T2	T2	T2
53	T2	T2	T2	T2	T2	T2
54	T2	T2	T2	T2	T2	T2
55	T2	T2	T2	T2	T2	T2
56	T1 y T2	T1 y T2	T2	T1	T2	T2
57	T2	T2	T2	T2	T2	T2
58	T2	T2	T2	T2	T2	T2
59	T2	T2	T2	T2	T2	T2
60	T1 y T2	T1 y T2	T2	T1	T2	T2
61	T1 y T2	T1 y T2	T2	T1	T2	T2
62	T1 y T2	T1 y T2	T2	T1	T2	T2
63	T2	T2	T2	T2	T2	T2
64	T2	T2	T2	T2	T2	T2
65	T3 y T4	T3 y T4	T3	T2	T3	T4
66	T3 y T4	T3 y T4	T4	T3	T3	T4
67	T3 y T4	T3 y T4	T3	T4	T3	T4
68	T3 y T4	T3 y T4	T4	T3	T3	T4
69	T3 y T4	T3 y T4	T3	T4	T3	T4
70	T3 y T4	T3 y T4	T4	T3	T3	T4
71	T3 y T4	T3 y T4	T3	T4	T3	T4
72	T3 y T4	T3 y T4	T4	T3	T3	T4
73	T3 y T4	T3 y T4	T3	T4	T3	T4
74	T3 y T4	T3 y T4	T4	T3	T3	T4
75	T3 y T4	T3 y T4	T3	T4	T3	T4
76	T3 y T4	T3 y T4	T4	T3	T4	T3
77	T3 y T4	T3 y T4	T3	T4	T4	T3
78	T3 y T4	T3 y T4	T4	T3	T4	T3
79	T3 y T4	T3 y T4	T3	T4	T4	T3
80	T3 y T4	T3 y T4	T4	T3	T4	T3
81	T3 y T4	T3 y T4	T3	T4	T4	T3
82	T3 y T4	T3 y T4	T4	T3	T4	T3
83	T3 y T4	T3 y T4	T3	T4	T4	T3
84	T3 y T4	T3 y T4	T4	T3	T4	T3
85	T3 y T4	T3 y T4	T3	T4	T4	T3
86	T3 y T4	T3 y T4	T4	T3	T4	T3
87	T3 y T4	T3 y T4	T3	T4	T4	T3
88		T3 y T4	T4	T3	T3	T4
89		T3 y T4	T3	T4	T3	T4
90		T3 y T4	T4	T3	T3	T4
91		T3 y T4	T3	T4	T3	T4
92		T3 y T4	T4	T3	T4	T3
93	T1	T1	T1	T1	T1	T1
94	T1	T1	T1	T1	T1	T1
95	T1	T1	T1	T1	T1	T1
96	T1	T1	T1	T1	T1	T1
97	T1	T1	T1	T1	T1	T1
98		T3 y T4	T4	T3	T4	T3
99		T3 y T4	T3	T4	T4	T3
100		T3 y T4	T4	T3	T4	T3

Anexo 2.8**Cuestionario sobre la rúbrica y el proceso para evaluar los textos del Examen de Español como Lengua Extranjera para el Ámbito Académico**

Cuestionario sobre la rúbrica y el proceso para evaluar los textos del Examen de Español como Lengua Extranjera para el Ámbito Académico

El siguiente formulario servirá para evaluar el diseño y descripción de las categorías de la rúbrica. La información será tratada de manera confidencial y se utilizará para conducir investigaciones relacionadas con la prueba escrita del examen.

*Obligatorio

1. ¿Cuál es su formación académica? *

Incluya diplomados, cursos y talleres en evaluación

2. ¿Ha dado clases de español a extranjeros? *

Si su respuesta es sí, ¿durante cuánto tiempo y a qué niveles?

3. ¿Qué otros idiomas habla y a qué nivel? ***4. ¿Había participado antes como evaluador de pruebas escritas? ***

Si su respuesta es sí, ¿En qué pruebas y durante cuánto tiempo?

5. ¿Cuál es su opinión general sobre la rúbrica? *

6. ¿Considera usted que hay alguna categoría que debería agregarse o suprimirse? *

Justifique su respuesta.

7. ¿Considera usted que el número de bandas (0 al 3) es suficiente para evaluar cada criterio? *

Justifique su respuesta.

8. ¿Hubo alguna sección de la rúbrica que le pareciera poco clara para evaluar los textos? *

Justifique su respuesta.

9. ¿Hubo alguna categoría o nivel que le fue particularmente difícil emplear? *

Justifique su respuesta

10. En caso de haber encontrado dificultades en el uso de alguna categoría o nivel de la rúbrica, ¿qué estrategia o criterio empleó para evaluar dicha categoría o nivel? *

11. Según su criterio, jerarquice las categorías de la rúbrica de mayor a menor importancia: *

Logro de la tarea, desarrollo de ideas, cohesión, coherencia, léxico, gramática y ortografía

12. Según su opinión, ¿considera que las 4 temáticas eran equiparables en dificultad? *

Justifique su respuesta.

13. ¿Encontró alguna diferencia sustancial al evaluar textos escritos a mano y en computadora? *

Justifique su respuesta.

14. Exprese su opinión de acuerdo con la retroalimentación que recibió en el uso de la rúbrica. *

15. ¿Considera usted que hizo falta algo en su proceso de evaluación de los textos? *

Si tiene algún comentario adicional, agradeceremos mucho nos lo haga saber.

Anexo 2.9

Resultados del cuestionario sobre la rúbrica y el proceso para evaluar los textos del EXELAA

1. ¿Cuál es su formación académica?

Evaluador 1	
Evaluador 2	Curso de formación de profesores como LE. Maestría en Lingüística Aplicada. Doctorando en Lingüística
Evaluador 3	Especialización en Literatura mexicana del siglo XX, Lic. en Diseño Gráfico, Pasante en Letras Hispánicas, Diplomado en Español como Lengua Adicional (CEPE), Diplomado de Español como Lengua Extranjera (International House).
Evaluador 4	Lic. en Pedagogía FFYL-UNAM Lic. en Enseñanza de Español L.E., FES-ACATLÁN, UNAM Especialización en Enseñanza de Español L.E. Diploma en Enseñanza de Francés L.E. Evaluación cursos interanuales e intersemestrales UNAM, Talleres, etc.
Evaluador 5	Licenciatura en filología inglesa.
Evaluador 6	Licenciatura en letras. Curso de formación de profesores como LE

2. ¿Ha dado clases de español a extranjeros?

Evaluador 1	
Evaluador 2	Sí, 10 años. Niveles básicos, intermedios, avanzados. Clases de redacción, conversación. Cursos de preparación para exámenes de certificación.
Evaluador 3	Sí, 8 años, Básico 1-3, Intermedio 1 y 2.
Evaluador 4	Si. Desde 2006 a la fecha. Todos los niveles
Evaluador 5	Sí. Durante 1 año. Niveles A1, A2 y B2
Evaluador 6	Si, 10 años, básico, intermedio y avanzado

3. ¿Qué otros idiomas habla y a qué nivel?

Evaluador 1	
Evaluador 2	Inglés C1, alemán B2, francés e italiano A1
Evaluador 3	Inglés intermedio, francés avanzado
Evaluador 4	inglés 100%, francés 95%, alemán 80%, catalán 60%
Evaluador 5	Inglés, Nivel C1, francés A2
Evaluador 6	Inglés avanzado

4. ¿Había participado antes como evaluador de pruebas escritas?

Evaluador 1	
Evaluador 2	Sí. EXELEAA
Evaluador 3	No.
Evaluador 4	Si. CENEVAL. EXPRESSE. 2014 y 2015. Evaluación de lenguas extranjeras en diferentes instituciones por alrededor de 35 años.
Evaluador 5	No.
Evaluador 6	sí, exámenes finales de aprovechamiento, 10 años

5. ¿Cuál es su opinión general sobre la rúbrica?

Evaluador 3	La rúbrica es clara, la descripción de cada competencia es fácil de comprender, un texto muestra para cada caso podría ser de utilidad
Evaluador 4	En general los modelos de rúbricas son adecuados para los fines que se persiguen. Una observación que haría sería presentarlas de manera más visual; es decir, que se compacte la información (sin eliminarla) en una o dos tablas. Es muy laborioso estar viendo en 4 o 5 páginas, todos los conceptos a evaluar para cada examen.
Evaluador 5	Me parece que en general era una buena rúbrica.
Evaluador 6	Los rubros despliegan una serie de opciones que permite tener una idea general de lo que se quiere evaluar. No hay ejemplos específicos, por lo que el evaluador requiere de cierta práctica para saber con claridad qué se está evaluando. En mi experiencia, evaluar exámenes de redacción en inglés y en español es diferente. El inglés se concentra en la precisión gramatical así como en el uso pertinente de los conectores discursivos, sin cambiar preposiciones ni sentido de los mismos. Situación diferente para el español. No es tan estricto el control de estos aspectos. Pero, puedo decir que de todo se aprende. Al final, son lenguas diferentes y con sus complejidades propias.

6. ¿Considera usted que hay alguna categoría que debería agregarse o suprimirse?

Evaluador 3	Creo que las categorías son adecuadas.
Evaluador 4	El número de palabras, si no se va a considerar no veo razón para conservarlo. En cuanto al rubro: En desarrollo, tal vez se podía usar otra terminología, puesto que se trata en general de alumnos con un nivel medio a avanzado de la lengua. Tal vez sería más adecuado usar los términos tradicionales de suficiente, incipiente o insuficiente.
Evaluador 5	No. Creo que la rúbrica era bastante precisa.
Evaluador 6	Añadir ejemplos de todos los rubros.

7. ¿Considera usted que el número de bandas (0 al 3) es suficiente para evaluar cada criterio?

Evaluador 3	Sí.
Evaluador 4	Es difícil que todos los elementos a evaluar se adecuen al mismo número de bandas. Tal vez cada criterio debería tener un diferente número, no obstante no ser uniforme con los otros (algunos podrían tener 5 o 4, de acuerdo con la especificidad requerida)
Evaluador 5	Sí, aunque también podría ser del 1 al 5 o del 1 al 10.
Evaluador 6	Pienso que no, hay detalles en la escritura que no entran en sólo 1 o 2 o 3. Esto también lo hizo interesante.

8. ¿Hubo alguna sección de la rúbrica que le pareciera poco clara para evaluar los textos?

Evaluador 3	La referente al "léxico especializado", como son temas de interés general se me dificultó un poco establecer el criterio de evaluación, sólo fue posible contrastando las pruebas de varios alumnos.
Evaluador 4	La cohesión y la coherencia pueden estar muy relacionadas. la coherencia de un texto se logra a través de la correcta organización y distribución de la información sobre un determinado tema, haciendo posible su comprensión. En el rubro excepcional o experto, no me queda muy clara la diferenciación entre cohesión y

	<p>coherencia. Aunque a la primera se le identifica como microestructura, y la segunda como microestructura, ambas hacen referencia a los nexos y marcadores discursivos en los párrafos.</p> <p>Por otro la ausencia de uso de marcadores discursivos no siempre denota una total falta de estructuración. Puede haber un orden y una lógica a pesar de todo. Los conceptos de excepcional y de experto (sobre todo el segundo), me parecen un poco difíciles de tener, aun cuando hay alumnos bastante familiarizados, me pareciera que “es mucho pedir” para una lengua extranjera.</p> <p>Es cómo el concepto tan discutido del “dominio” de una lengua”. Me parece mejor usar “nivel” como en la Pregunta 3</p>
Evaluador 5	Para mí todo estaba bastante claro.
Evaluador 6	No

9. ¿Hubo alguna categoría o nivel que le fue particularmente difícil emplear?

Evaluador 3	No
Evaluador 4	La anterior por lo expuesto.
Evaluador 5	No, todo estaba bien explicado.
Evaluador 6	En general, todas las categorías, porque, por ejemplo, si el texto usaba 1 o 2 palabras que se considera léxico adecuado, pero el texto era largo, no sabía si esas 2 palabras eran suficientes para darle un puntaje alto o no. Y, lo mismo sucede con el resto de las categorías.

10. En caso de haber encontrado dificultades en el uso de alguna categoría o nivel de la rúbrica, ¿qué estrategia o criterio empleó para evaluar dicha categoría o nivel?

Evaluador 3	Por ejemplo, cuando un texto era mucho más largo de lo solicitado (la rúbrica señala hasta 20% más) pero en ocasiones doblaba la cantidad de palabras, pero la evaluación general del texto era buena, ignoré la cantidad de palabras. Es cierto, que poder resumir es importante pero ante un texto claro y bien argumentado creo que ya no lo es tanto. Creo que siempre antepuse el hecho de que se cumpliera el objetivo comunicativo.
Evaluador 4	Buscar la equivalencia con términos como suficiente, insuficiente y otros.
Evaluador 5	No encontré dificultades.
Evaluador 6	Utilicé el mínimo uso de... Si aparecía, como mencioné antes, 1 o 2 palabras que entraran en tal o cual categoría, así asignaba los puntajes.

11. Según su criterio, jerarquice las categorías de la rúbrica de mayor a menor importancia.

Evaluador 1	
Evaluador 2	La adecuación o el logro de la tarea y el desarrollo de las ideas. En segundo lugar, la coherencia y después la cohesión. Por último el uso del léxico y finalmente, la gramática y también la ortografía.
Evaluador 3	Logro de la tarea, desarrollo de ideas, cohesión y coherencia, gramática, léxico y ortografía.
Evaluador 4	<p>La jerarquización debe hacerse, considero, de acuerdo con una necesidad específica de la lengua. Para estudiar un posgrado por ejemplo: Gramática y ortografía, y léxico en primer lugar puesto que nos dan el desarrollo de ideas, el logro de tareas y la cohesión y coherencia.</p> <p>Y para fines comunicativos: Léxico - Desarrollo de ideas - Logro de la tarea - Coherencia y cohesión - gramática y ortografía.</p>

Evaluador 5	Logro de tarea, cohesión, coherencia, desarrollo de ideas, gramática, léxico y ortografía.
Evaluador 6	Gramática, ortografía, cohesión, coherencia, léxico, desarrollo de ideas, logro de la tarea. Éste último no existe si no hay relación estrecha entre los demás rubros, según yo. Pero, siempre hay aprendizaje en todo.

12. Según su opinión, ¿considera que las 4 temáticas eran equiparables en dificultad?

Evaluador 3	Creo que la dificultad es similar, sin embargo, en las pruebas que calificué pareciera que fue más inspirador el tema de la salud, porque en general creo que desarrollaron más ideas.
Evaluador 4	No, por eso es necesario tener diferentes bandas. Si por ejemplo, faltan acentos en exceso, se puede tener 0 pero todo lo demás ser muy bien hecho.
Evaluador 5	No. Las gráficas eran más difíciles que los ensayos y algunos alumnos tendían a confundirse al no poder interpretarlas adecuadamente.
Evaluador 6	Sí

13. ¿Encontró alguna diferencia sustancial al evaluar textos escritos a mano y en computadora?

Evaluador 3	Es más fácil en la computadora, porque son más legibles los textos, aunque la separación de párrafos se perdió.
Evaluador 4	No realmente, estoy acostumbrada a todo tipo de letra y de cualquier manera me era necesario tener las pruebas impresas.
Evaluador 5	Sí. Era más fácil leer los textos escritos en computadora.
Evaluador 6	Lo inteligible de los textos. A mano era complicado, a veces, entender las palabras; en la computadora es mucho más fácil leer

14. Exprese su opinión de acuerdo con la retroalimentación que recibió en el uso de la rúbrica.

Evaluador 3	Fue muy importante para aclarar dudas y unificar criterios.
Evaluador 4	No me quedó muy clara.
Evaluador 5	Fue precisa y me ayudó a emplear la rúbrica lo mejor posible.
Evaluador 6	Como mencioné antes, yo tenía una concepción de evaluar de un modo, pero al recibir cierta explicación para el uso de la rúbrica, me di cuenta de los criterios utilizados para evaluar estos textos. Fue un enfoque diferente y, aprendí algo nuevo, pero, sin duda, es necesaria mayor exposición a este tipo de criterios.

15. ¿Considera usted que hizo falta algo en su proceso de evaluación de los textos?

Evaluador 3	No
Evaluador 4	Seguramente no hizo falta, pero tal vez hay que enfatizar que se trata de una lengua extranjera.
Evaluador 5	Creo que no. Todo estuvo bien.
Evaluador 6	Capacitación para usar la rúbrica.

Si tiene algún comentario adicional, agradeceremos mucho nos lo haga saber.

Evaluador 3	El proceso fue muy interesante, para mí que no tenía experiencia en este tipo de evaluación (salvo calificar exámenes en cursos convencionales), fue de gran ayuda. Aprendí a aplicar baremos para hacer más objetiva la evaluación.
Evaluador 4	Fue una experiencia interesante, sobre todo conocer esta línea de investigación, la cual además es una labor cotidiana del docente de L.E.
Evaluador 5	
Evaluador 6	

Anexo 2.10
Habilidades de escritura que mide la segunda versión de la rúbrica

Habilidades de escritura	EXPERTO					
	1	2	3	4	5	
20. Dar crédito adecuado a las fuentes.	0	0	0	1	0	X
23. Organizar las ideas y la información coherentemente.	1	1	1	1	1	Sí
33. Evitar errores ortográficos.	1	1	1	1	1	Sí
24. Escribir de forma precisa y concisa, evitando frases vagas o vacías.	1	1	1	1	0	Sí
3. Abstractar o resumir información esencial.	1	1	0	0	1	Sí
16. Presentar los datos y la información de manera clara y lógica.	1	1	1	1	1	Sí
6. Analizar y sintetizar información de múltiples fuentes (incluye comparación y contraste).	1	1	0	1	0	Sí
5. Explicar un hecho o suceso utilizando evidencia.	1	0	0	1	1	Sí
29. Revisar y editar el texto para mejorar su calidad, coherencia y precisión.	1	0	0	0	1	X
19. Integrar el material citado y referenciado debidamente dentro del texto.	1	0	0	1	0	X
26. Escribir con fluidez, evitando que el lenguaje sea complicado y rebuscado.	0	0	1	0	1	X
32. Utilizar correctamente la gramática del español escrito con el fin de evitar que el lector se distraiga o que se cometan errores que alteren el significado.	1	1	1	1	1	Sí
12. Interpretar los datos dentro de un marco determinado.	0	1	0	0	0	X
8. Escribir persuasivamente mediante la construcción de un argumento bien razonado para apoyar o refutar una postura.	1	1	1	1	1	Sí
30. Usar vocabulario técnico y específico al contenido del tema con precisión y de forma apropiada para un propósito y audiencia en particular.	1	1	1	1	1	Sí
10. Examinar el razonamiento en un argumento dado y discutir las fortalezas y debilidades de su lógica.	1	1	0	0	1	Sí
9. Explorar las relaciones entre ideas complejas y posiblemente contradictorias.	1	1	1	0	1	Sí
31. Elegir palabras efectivamente.	1	1	1	1	1	Sí
2. Explicar cómo se lleva a cabo un procedimiento.	1	0	0	1	0	X
15. Escribir adecuadamente para un público bien informado y reflexivo.	1	1	1	1	1	Sí
21. Desarrollar un debate bien enfocado y debidamente sustentado mediante ejemplos relevantes.	1	0	1	1	1	Sí
11. Identificar los problemas en una línea de acción propuesta o en la interpretación de hechos y proponer soluciones o alternativas de interpretación.	0	1	0	0	0	X
1. Describir observaciones.	0	1	0	0	1	X
25. Escribir con claridad mediante transiciones suaves de un pensamiento a otro.	0	1	1	1	1	Sí
4. Expresar opiniones personales.	1	1	1	1	1	Sí
22. Aclarar las relaciones entre ideas principales y secundarias	1	1	1	1	1	Sí
13. Clasificar la información de acuerdo con categorías o jerarquías.	0	1	0	1	0	X
7. Predecir consecuencias o resultados mediante el análisis de información, patrones o procesos.	1	0	0	0	0	X
27. Variar la estructura de las oraciones para comunicar ideas de manera efectiva.	1	0	1	0	0	X
18. Utilizar formatos eficientes y claros para organizar la información y guiar al lector.	1	0	1	0	1	Sí
14. Utilizar las convenciones de un género en particular.	1	1	1	1	1	Sí
28. Expresar ideas en forma original o novedosa para mantener el interés del lector.	0	1	0	1	0	X

ANEXOS ESTUDIO 3

Anexo 3.1

Estudiantes de posgrado de la UNAM que participaron en las entrevistas sobre escritura académica

Número de examinado	Edad	Programa de estudios	Nacionalidad
4	41	Maestría en Lingüística aplicada	Canadiense
6	26	Doctorado en Ciencias de la Tierra	Belga
8	25	Maestría en Relaciones Internacionales	Checa
15	29	Doctorado en Ciencias Políticas y Sociales	Francesa
16	Sin dato	Postdoctorado en Ciencias de la Tierra	Belga
20	30	Doctorado en Lingüística	Austriaca
66	40	Doctorado en ciencias políticas y sociales	Polaca
67	27	Doctorado en ciencias de la tierra	Belga
68	29	Doctorado en letras	Alemana
69	24	Maestría en Estudios Latinoamericanos	Alemana
70	28	Maestría en Historia del Arte	Checa
72	Sin dato	Doctorado en estudios Latinoamericanos	Trinidad y Tobago
73	21	Maestría en Geografía	Alemana
74	24	Maestría en Estudios Latinoamericanos	Italiana
75	21	Maestría en Geografía ambiental	Alemana
76	22	Doctorado en Ciencias Políticas	China
77	32	Doctorado en Historia	Surcoreana
78	28	Maestría en Estudios Mesoamericanos	Italiana
79	30	Maestría en Artes Visuales	Estadounidense
80	25	Maestría en Estudios Mesoamericanos	China
82	29	Doctorado en Estudios Mesoamericanos	Alemana
83	29	Maestría en Administración	Estadounidense
84	23	Maestría en Estudios Latinoamericanos	Italiana
85	34	Doctorado en Artes y Diseño	Japonesa
86	24	Maestría en Estudios Latinoamericanos	Inglesa

Anexo 3.2

Encuesta sobre el dominio de las habilidades de escritura académica en el posgrado

LA ESCRITURA ACADÉMICA EN EL POSGRADO DE LA UNAM

El siguiente cuestionario busca identificar la importancia de la escritura académica en los diversos programas del Posgrado de la UNAM. La información que se recabará es confidencial y podrás solicitar que se te envíen los resultados del estudio. El cuestionario consta de dos partes; en la primera, aparece una serie de preguntas abiertas sobre tu perfil, mientras que en la segunda, se te pregunta sobre las habilidades de escritura que posees.

De antemano, agradecemos tu valiosa participación.

***Obligatorio**

DATOS PERSONALES

Nombre completo *

Lengua materna *

¿Cuánto tiempo tienes en México? *

¿En qué maestría o doctorado estás registrado? *

¿Qué semestre de maestría o doctorado cursas?

¿Qué materias cursas?

Continuar »

Anexo 3.3

Guion de entrevista semiestructurada para estudiantes de posgrado no hispanohablantes de la UNAM

OBJETIVO: Investigar cuáles son las habilidades y dificultades que presentan los estudiantes de posgrado no hispanohablantes al momento de escribir en el contexto universitario, así como los factores contextuales asociados a éstas.

PERCEPCIÓN SOBRE EL DOMINIO DE LA ESCRITURA ACADÉMICA

¿Cuánto tiempo tienes escribiendo en español en un contexto universitario?

¿Cómo consideras que es tu nivel de escritura en la universidad? (por ejemplo: malo, suficiente, bueno, muy bueno, excelente)

¿Cómo te sientes escribiendo en español? (por ejemplo: frustrado, con temor, con desconfianza, con confianza, emocionado, seguro)

¿Cuánto consideras que te tardas en escribir en español? (por ejemplo: lo normal, más de lo normal, mucho más de lo normal, muchísimo más de lo normal)

FACTORES CONTEXTUALES QUE FACILITAN O DIFICULTAN LA ESCRITURA ACADÉMICA

¿Tomas clases en este momento?

¿En qué idioma está el material que lees normalmente?

¿En qué idioma tomas tus notas? ¿Son a mano a computadora?

¿Tu lengua materna, facilita o dificulta tu proceso de escritura en español?

LA ESCRITURA EN EL CONTEXTO ACADÉMICO

EXÁMENES

¿Tienes o has tenido exámenes escritos en clase?

¿Qué dificultades encuentras al momento de escribir un examen en clase?

TRABAJOS

¿Te han evaluado con trabajos escritos?

¿Cuál consideras que es tu mayor dificultad al escribir en español? (por ejemplo: gramática, nexos, vocabulario, coherencia, cohesión, tipología textual).

¿Qué trabajo académico te ha costado más escribir y por qué? (por ejemplo: hacer resumen de artículo o libro, abstract, ensayo, monografía, resumen de artículo en inglés)

¿Cómo es tu proceso de escritura?

RECURSOS EMPLEADOS DURANTE EL PROCESO DE ESCRITURA ACADÉMICA

¿Qué recursos empleas cuando escribes? (por ejemplo: diccionario, tesauros, traductores, pides ayuda a alguien).

¿Qué tipo de retroalimentación recibes de tus profesores? ¿Es suficiente?

¿Qué crees que te ayudaría a mejorar tu escritura académica?

Anexo 3.4

Rubrica para evaluar las entrevistas sobre la habilidad de escritura académica de los estudiantes no hispanohablantes del posgrado de la UNAM

Según lo que escucha usted en la grabación, registre la información que proporcionan los candidatos respecto a los siguientes rubros:	
¿Cuánto tiempo tiene escribiendo en un contexto académico en español?	
Menos de un mes	0
Entre 1 y 6 meses	1
Entre 6 meses y 1 año	2
Más de un año	3
No hay suficiente información para evaluar su desempeño en este rubro.	NA
¿Qué sensaciones describe al momento de escribir en español?	
Muy negativas	0
Negativas	1
Positivas o neutras	2
Muy positivas, aunque reconoce que puede mejorar.	3
No hay suficiente información para evaluar su desempeño en este rubro	NA
¿Cuánto tiempo requiere para escribir en español?	
Aún no escribe en un contexto académico o el tiempo que le toma escribir es excesivamente mayor al que le tomaría hacerlo en su lengua materna.	0
El tiempo que le toma escribir en español es considerablemente mayor al que le toma hacerlo en lengua materna (el doble).	1
El tiempo que le toma escribir en español es un poco más que lo que le toma en lengua materna.	2
El tiempo que le toma escribir es prácticamente igual que en lengua materna o en la lengua en que está acostumbrado a escribir.	3
No hay suficiente información para evaluar su desempeño en este rubro	NA
¿Ha tomado o toma clases en su posgrado?	
Ninguna	0
Por lo menos una	1
Dos o tres	2
Más de tres	3
No hay suficiente información para evaluar su desempeño en este rubro	NA
¿Toma clase de lengua o considera que las necesita?	
Toma clases de lengua.	0
No toma clases de lengua y reconoce que las necesita.	1
No toma clases, aunque considera que le ayudarían cursos de escritura.	2
No toma clases y considera que no las necesita.	3
No hay suficiente información para evaluar su desempeño en este rubro	NA
¿Aproximadamente qué cantidad de material leído por el candidato en sus cursos de posgrado está en español?	
Ningún material (sólo se leen textos en otro idioma).	0
La mitad.	1
Más de la mitad.	2
Prácticamente todo.	3
No hay suficiente información para evaluar su desempeño en este rubro	NA

¿Si el candidato hace lecturas extras, ¿opta por hacerlas en español?	
Nunca, porque considera su nivel de lengua insuficiente.	0
En pocas ocasiones, porque no existe material pertinente para su investigación o porque prefiere lecturas en lengua extranjera.	1
En varias ocasiones, dependiendo de la pertinencia para su investigación. No le importa leer en una u otra lengua.	2
En múltiples ocasiones, porque las utiliza como estrategia para favorecer su escritura. Prefiere lecturas en español.	3
No hay suficiente información para evaluar su desempeño en este rubro	NA
¿El candidato utiliza el español para tomar notas en la clase?	
Nunca, escribe la mayoría de sus notas en lengua materna o en alguna otra.	0
En pocas ocasiones, pero más bien mezcla sus notas en lengua materna y español.	1
En varias ocasiones, salvo cuando no comprende conceptos importantes en español o no sabe cómo traducirlos.	2
En múltiples ocasiones.	3
No hay suficiente información para evaluar su desempeño en este rubro	NA
¿Qué experiencia tiene escribiendo en un contexto académico (trabajos y exámenes)?	
Ninguna	0
Tiene muy poca experiencia con trabajos y exámenes escritos en español.	1
Ha escrito algunos trabajos académicos y/o quizás también exámenes.	2
Se encuentra ampliamente familiarizado con los trabajos y/o exámenes escritos en el ámbito universitario.	3
No hay suficiente información para evaluar su desempeño en este rubro	NA
¿Cómo describe el candidato su proceso de escritura en español?	
Insuficiente para producir textos en español.	0
Poco eficiente porque se basa en la traducción de la lengua materna al español.	1
Adecuado, aunque los huecos léxicos, gramaticales o sintácticos en ocasiones entorpecen el flujo de sus ideas.	2
Fluido, aunque con algunos titubeos en la gramática o vocabulario que prácticamente no impiden el flujo de sus ideas.	3
No hay suficiente información para evaluar su desempeño en este rubro	NA
¿Con qué frecuencia se apoya en recursos materiales (traductores, diccionarios o foros) para elaborar sus textos?	
Siempre y sin estrategia alguna, pues su nivel de lengua no es suficiente para utilizarlos de manera adecuada.	0
En múltiples ocasiones y con pocas estrategias de uso.	1
En varias ocasiones, pero con algunas dificultades de uso.	2
Esporádicamente (porque no los requiere) y los usa de manera eficiente	3
No hay suficiente información para evaluar su desempeño en este rubro	NA
¿Cuenta con lectores (amigos, tutor, compañeros, etc.) que retroalimenten sus trabajos escritos y le permitan mejorar así su escritura?	
Con ninguno, porque no cuenta con estrategias de apoyo.	0
Con pocos y la retroalimentación no es suficiente.	1
Con varios lectores que le permiten mejorar su escritura.	2
Con diversos lectores. Cuenta con sofisticadas estrategias de apoyo dependiendo del tipo de retroalimentación que necesita.	3
No hay suficiente información para evaluar su desempeño en este rubro	NA

Anexo 3.5

Niveles de logro de la prueba escrita del Examen de Español para el Ámbito Académico

EXPRESIÓN ESCRITA	
Nivel bajo (menor a B1) De 0 a 1.275 puntos	El candidato no cuenta con el nivel suficiente producir textos académicos de tipo argumentativo. Su capacidad de escritura se limita a escribir textos básicos sobre la vida cotidiana.
Nivel intermedio bajo (B1) De 1.276 a 1.875 puntos	<p>El candidato es capaz de escribir textos sencillos relacionados con el ámbito académico. El candidato es capaz de</p> <ul style="list-style-type: none"> • Emplear vocabulario académico básico y repetitivo. • Expresar las ideas principales de forma elemental, y enlazadas de forma básica. • Escribir textos argumentativos sencillos y breves, cohesionados de forma elemental en una secuencia lineal. Comunica y ofrece su opinión de manera sencilla. • Describir y contrastar información de forma básica con un grado elemental de análisis. • Disponer de elementos lingüísticos limitados para describir o explicar los puntos principales de una idea o un problema. • Expresarse con un buen control gramatical, aunque con una influencia evidente de la lengua materna. Comete errores, pero queda claro lo que intenta expresar. • Producir una escritura continua que suele ser inteligible en toda su extensión, aunque con recurrentes errores ortográficos o de puntuación.
Nivel intermedio alto (B2) De 1.876 a 2.475 puntos	<p>El candidato es capaz de escribir textos claros y detallados relacionados con el ámbito académico. El candidato es capaz de:</p> <ul style="list-style-type: none"> • Escribir textos argumentativos razonando a favor o en contra de un punto de vista concreto y explicando las ventajas y las desventajas de varias opciones. • Describir y contrastar información con detalle, aunque con algunas imprecisiones. • Emplear suficientes elementos lingüísticos como para hacer descripciones claras, expresar puntos de vista y desarrollar argumentos, utilizando para ello algunas oraciones complejas. • Utilizar vocabulario académico variado y preciso, aunque no especializado. Su precisión léxica es generalmente alta, aunque tenga alguna confusión o cometa alguna incorrección al seleccionar las palabras, sin que ello obstaculice la comunicación. • Demostrar un alto control gramatical, aunque todavía comete deslices esporádicos, errores no sistemáticos y pequeños fallos en la estructura de la frase, pero son escasos y a menudo puede corregirlos retrospectivamente. No comete errores que produzcan mal entendidos. • Producir una escritura continua inteligible que sigue las convenciones de organización y de distribución en párrafos. • Utilizar la ortografía y la puntuación de forma correcta, pero puede manifestar la influencia de la lengua materna. • Utilizar con eficacia, aunque no siempre, una variedad de palabras de enlace para señalar con claridad las relaciones que existen entre las ideas.

<p>Nivel superior (C1 o superior)</p> <p>De 2.476 a 3 puntos</p>	<p>El candidato es capaz de expresarse en textos claros y bien estructurados sobre temas complejos, resaltando lo aspectos que considera importantes. Selecciona el estilo apropiado para los lectores a quienes va dirigido el texto. El candidato es capaz de:</p> <ul style="list-style-type: none">• Escribir textos argumentativos claros y bien estructurados sobre temas complejos resaltando las ideas principales, ampliando con cierta extensión y defendiendo sus puntos de vista con ideas complementarias, motivos y ejemplos adecuados, y terminando con una conclusión apropiada.• Precisar sus opiniones y afirmaciones aportando grados de, por ejemplo, certeza/incertidumbre, creencia/duda, probabilidad, etc.• Describir y contrastar información con gran precisión y detalle.• Elegir dentro de una amplia gama de formulaciones gramaticales y sintácticas para expresarse con claridad y sin tener que limitar lo que quiere decir.• Emplear un amplio repertorio léxico que le permite superar con soltura sus deficiencias mediante circunloquios; apenas se le nota que busca expresiones o que utiliza estrategias de evitación. Pequeños y esporádicos deslices, pero sin errores importantes de vocabulario.• Mantener un alto grado de corrección gramatical de modo consistente; los errores son escasos y apenas se notan.• Emplear la estructura, la distribución en párrafos y la puntuación de forma apropiada.• Usar correctamente la ortografía, salvo deslices tipográficos de carácter esporádico.• Producir un discurso claro, fluido y bien estructurado, mostrando un uso adecuado de criterios de organización, conectores y mecanismos de cohesión.
--	--