



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

DOCTORADO EN LINGÜÍSTICA

**LA RESPONSABILIDAD PROFESIONAL EN EL DISCURSO SOBRE LA PRÁCTICA
EVALUATIVA DEL PROFESOR DE LENGUA EXTRANJERA EN EDUCACIÓN
SUPERIOR**

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN LINGÜÍSTICA

PRESENTA:

HUGO ABELARDO ANDRADE MAYER

TUTORA PRINCIPAL

**DRA. LAURA GABRIELA GARCÍA LANDA.
ESCUELA NACIONAL DE LENGUAS, LINGÜÍSTICA Y TRADUCCIÓN**

MIEMBROS DEL COMITÉ TUTOR:

**DRA. MA. DEL CARMEN CONTIJOCH ESCONTRIA.
ESCUELA NACIONAL DE LENGUAS, LINGÜÍSTICA Y TRADUCCIÓN**

**DRA. NOËLLE GROULT BOIS.
ESCUELA NACIONAL DE LENGUAS, LINGÜÍSTICA Y TRADUCCIÓN**

**DRA. BÉATRICE FLORENCE BLIN.
ESCUELA NACIONAL DE LENGUAS, LINGÜÍSTICA Y TRADUCCIÓN**

**DRA. PAULINE MARION DOROTHY MOORE.
UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO**

CIUDAD UNIVERSITARIA, CIUDAD DE MÉXICO

ABRIL, 2023



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.

AGRADECIMIENTOS

A Dios. Toda la Gloria a Él.

A mis padres Hugo y Susana y mi hermana Susana.

A Laura García Landa, por haber confiado y apoyado este proyecto. Por su dedicación, paciencia, interés, disposición y tiempo.

A Carmen Contijoch, Noëlle Groult, Béatrice Blin y Pauline Moore, mi comité tutorial, por sus aportaciones e interés en el proyecto.

A Pauline, Virna, Alfredo y Georgia, egresados de este mismo posgrado, que me hicieron ver que era la mejor opción para mí y, no se equivocaron, así fue.

A Virna, por sus consejos, apoyo y ánimos permanentes.

A Pauline, que ha estado presente de manera muy cercana en mi formación, por su confianza y ayuda incondicional.

A Manolo, Pilar y Alfredo, que en distintos grados contribuyeron con su ayuda a este logro.

A Bruno.

A Nadia, Chío, Jera, Gaby, Flor, Yare, Dale, Fer, María, Dany y Pita, por su trabajo y paciencia.

“Los galaaditas ocuparon los vados del Jordán que conducen a Efraín, y cada vez que algún sobreviviente de Efraín decía: «Déjenme cruzar», los hombres de Galaad le preguntaban: «¿Eres de la tribu de Efraín?» Si él contestaba: «No», ellos decían: «Muy bien, di “Shibolet”». Si decía: «Sibolet», porque no podía pronunciar la palabra correctamente, lo agarraban y allí mismo, en los vados del Jordán, lo degollaban. En aquella ocasión murieron cuarenta y dos mil hombres de la tribu de Efraín.”

Jueces 12:5-6 NVI

ÍNDICE

INTRODUCCIÓN	1
CAPÍTULO I.....	7
EVALUACIÓN EDUCATIVA Y DE LENGUAS	7
Introducción al capítulo.....	7
1. Evaluación	8
1.1 Evaluación desde la perspectiva de la complejidad.....	10
1.2 Evaluación de aprendizaje de lenguas.....	11
1.2.1 Enfoques en evaluación de lenguas.....	11
1.2.2 Evaluación de lengua en el salón de clase.....	15
1.2.2.1 Evaluación formativa y sumativa	16
1.2.2.2 Evaluación y el enfoque comunicativo	18
1.2.2.3 Prácticas evaluativas en el salón de clase	19
1.2.3 Literacidad en evaluación.....	22
1.2.3.1 Literacidad en evaluación de lenguas.....	22
1.2.3.2 Alcance de la literacidad en evaluación de profesores de lengua.....	26
1.3. Exámenes de lengua	29
1.3.1 Antecedentes de estudios de diseño de exámenes de lengua	30
1.3.2 Definición y características de los exámenes de lengua	32
1.3.3 Utilidad de un examen de lengua (<i>Test usefulness</i>).....	33
1.3.3.1 Validez.....	34
1.3.3.2 Confiabilidad	36
1.3.3.3 Practicidad	39
1.3.3.4 Autenticidad	40
1.3.3.5 Interactividad	41
1.3.3.6 Impacto	41
Conclusión al capítulo.....	42
CAPÍTULO II.....	44
ÉTICA EN LA EDUCACIÓN Y EVALUACIÓN DE LENGUAS	44
Introducción al capítulo.....	44
2.1 Ética	44
2.1.1 Concepto de ética	45

2.2 Ética en la educación	47
2.2.1 Alteridad.....	47
2.2.2 El docente universitario	48
2.2.3 Las actitudes del docente	50
2.2.4 El docente y la evaluación educativa.....	51
2.3 La ética en el diseño de exámenes de lengua	53
2.3.1 Enfoques de la ética en el diseño de exámenes de lengua	56
2.3.1.1 Enfoque en la validez.....	58
2.3.1.2 Enfoque en la equidad (<i>fairness</i>).....	58
Conclusión al capítulo	60
CAPÍTULO III	62
METODOLOGÍA DE INVESTIGACIÓN	62
Introducción al capítulo	62
3.1 Fundamentación epistemológica, ontológica y axiológica	63
3.1.1 La teoría de la complejidad como fundamento epistemológico	63
3.1.2 Fundamentación ontológica	65
3.1.3 Fundamentación axiológica	66
3.2 Planteamiento del problema	66
3.3 Preguntas de investigación	69
3.4 Objetivos	70
3.4.1 Objetivo general.....	70
3.4.2 Objetivos particulares	70
3.5 Enfoque metodológico	71
3.5.1 Corte y diseño de investigación	71
3.5.1.1 Corte de investigación	71
3.5.1.2 Diseños de investigación.....	73
3.5.2 Estrategia metodológica.....	75
3.5.2.1 Estudio de caso.....	75
3.5.2.2 Teoría fundamentada	75
3.6 Participantes	77
3.6.1 Sujetos en fase cuantitativa.	77
3.6.1.1 Edad.....	77
3.6.1.2 Género.....	78
3.6.1.3 Años de experiencia docente	79

3.6.1.4 Grado de estudios	80
3.6.1.5 Sujetos por organismo académico	81
3.6.2. Participantes en la entrevista.....	83
3.7 Instrumentos de recolección de información	84
3.7.1 Instrumento de recolección de información cuantitativa: Encuesta	84
3.7.2 Instrumentación: Guion de entrevista semiestructurada.	87
3.8 Métodos de recopilación y análisis de información.....	89
3.8.1 Recopilación de datos cuantitativos: Encuesta	91
3.8.2 Análisis de datos cuantitativos	92
3.8.2.1 Análisis factorial exploratorio	92
3.8.3 Recolección de datos: Entrevista.....	94
3.8.4 Análisis de datos cualitativos	94
3.8.4.1 Uso de software para el análisis de la información cualitativa	94
3.8.4.2 Codificación abierta.....	95
3.8.4.3 Codificación axial	97
3.8.4.4 Codificación selectiva	100
3.9 Rigor metodológico.....	101
CAPÍTULO IV	103
ANÁLISIS DE RESULTADOS	103
Introducción al capítulo.....	103
4.1 Fase cuantitativa: Encuesta.....	104
4.1.1 Análisis de datos	106
4.1.1.1 Análisis de estadística descriptiva.....	106
4.1.1.2 Análisis Factorial exploratorio	108
4.2 Fase cualitativa: Entrevistas	115
4.2.1 Codificación abierta	115
4.2.1.1 Relaciones entre códigos.....	117
4.2.2 Codificación axial	120
4.2.2.1 Grupos de códigos por dimensión.....	120
4.2.2.1.1 Dimensión lingüística	121
4.2.2.1.2 Dimensión técnica.....	122
4.2.2.1.3 Dimensión ética	123
4.2.2.1.4 Dimensión personal	124
4.2.2.1.5 Dimensión docente	124

4.2.3 Codificación selectiva.....	130
4.3 Análisis del discurso de los participantes	131
4.3.1 La dimensión lingüística en el discurso de los profesores	131
4.3.1.1 Habilidades, Gramática y Vocabulario	131
4.3.1.2 Temática.....	133
4.3.1.3 Nivel de lengua.....	134
4.3.1.4 Conocimiento.....	134
4.3.2 La dimensión técnica en el discurso de los profesores.....	135
4.3.2.1 Confiabilidad	136
4.3.2.2 Validez.....	140
4.3.2.3 Practicidad	145
4.3.2.4 Dificultad.....	146
4.3.2.5 Adaptación.....	147
4.3.2.6 Preparación y conocimiento técnico.....	149
4.3.2.7 Tecnicismo	150
4.3.2.8 Colaboración.....	151
4.3.2.9 Calificación	153
4.3.3 La dimensión ética en el discurso de los profesores.....	154
4.3.3.1 Subdimensión interpersonal	154
4.3.3.1.1 Alteridad.....	154
4.3.3.1.2 Cuidado del alumno	157
4.3.3.1.3 Exclusión e inclusión.....	159
4.3.3.1.4 Reflexión.....	167
4.3.3.1.5 Impacto	170
4.3.3.1.6 Diálogo	173
4.3.3.2 Subdimensión valorativa	174
4.3.3.2.1 Autoevaluación.....	174
4.3.3.2.2 Juicio sobre el alumno	175
4.3.3.2.3 Valoración.....	177
4.3.3.2.4 Obligación	182
4.3.3.2.5 Justicia	184
4.3.3.2.6 Importancia	185
4.3.4 La dimensión personal	188
4.3.4.1 Opinión	189

4.3.5 La dimensión docente	191
4.3.5.1 Evaluación alternativa y evaluación continua.....	191
4.3.5.2 Libro de texto	192
4.3.5.3 Política institucional.....	194
4.4 Análisis del discurso por temas	196
4.4.1 Proceso de diseño de exámenes.....	196
4.4.2 Características de los exámenes de lengua	198
4.4.3 Los exámenes como instrumentos de evaluación del aprendizaje de lenguas	199
4.4.4 Objetivo de los exámenes	201
4.4.5 Dudas con respecto a los exámenes de lengua.....	202
4.4.6. Autopercepción del profesor con respecto a la evaluación.....	204
4.4.7 Percepción de sus propios exámenes.....	205
4.4.8 Reflexión sobre el impacto de sus instrumentos en los alumnos.....	206
Conclusiones a la fase cualitativa de análisis	207
DISCUSIÓN DE RESULTADOS	209
5.1 Triangulación de resultados	211
5.2 Los resultados vistos desde la perspectiva de la complejidad.....	215
5.2.1 La evaluación del aprendizaje de lengua como un sistema.....	218
5.3 La responsabilidad profesional del profesor en la evaluación del aprendizaje de lengua	220
5.4 Profesionalismo y aspectos éticos de la práctica evaluativa	222
5.4.1 Ética interpersonal y valorativa	223
5.4.2 La alteridad.....	226
5.5 Las prácticas evaluativas de los profesores de la Universidad.....	228
5.5.1 El proceso de diseño de exámenes	228
5.5.2 Características de un examen de lengua	230
5.5.3 Las prácticas evaluativas y el logro de validez y confiabilidad	231
5.5.4 Percepciones de los exámenes como instrumentos de evaluación	231
5.6 Impacto del estudio en la literacidad de evaluación de lenguas.....	233
5.7 El carácter social de la evaluación en el salón de clase	235
5.8 Cambio en el paradigma educativo sobre la evaluación de lenguas.....	236
CONCLUSIONES.....	238
LISTA DE REFERENCIAS	242

ANEXOS	252
ANEXO A: REPORTE DE CÓDIGOS.....	252
ANEXO B: TABLA DE CO-OCURRENCIAS.....	252
ANEXO C: MATRIZ CÓDIGO-PARTICIPANTE	252
ANEXO D: REPORTE DE CÓDIGOS POR DIMENSIÓN.....	252
ANEXO E: VÍNCULOS CÓDIGO-CÓDIGO	252
ANEXO F: LISTA DE CÓDIGOS Y DESCRIPCIONES	253
ANEXO G: TRANSCRIPCIÓN DE ENTREVISTA DE LORENA	253
ANEXO H: TRANSCRIPCIÓN DE ENTREVISTA DE GENARO	253
ANEXO I: TRANSCRIPCIÓN DE ENTREVISTA DE GRACIELA.....	253
ANEXO J: TRANSCRIPCIÓN DE ENTREVISTA DE VIRGINIA	253
ANEXO K: TRANSCRIPCIÓN DE ENTREVISTA DE CAMILA.....	253
ANEXO L: TRANSCRIPCIÓN DE ENTREVISTA DE ALONDRA	254
ANEXO M: TRANSCRIPCIÓN DE ENTREVISTA DE BELINDA.....	254
ANEXO N: TRANSCRIPCIÓN DE ENTREVISTA DE SIMÓN	254
ANEXO O: TRANSCRIPCIÓN DE ENTREVISTA DE ERIKA.....	254
ANEXO P: TRANSCRIPCIÓN DE ENTREVISTA DE NORA	254
ANEXO Q: TRANSCRIPCIÓN DE ENTREVISTA DE GINA	254
ANEXO R: TRANSCRIPCIÓN DE ENTREVISTA DE NADIA	255
ANEXO S: CONCENTRADO DE CITAS DE ENTREVISTAS	255
ANEXO T: MATRIZ DE FACTOR ROTADO	256
ANEXO U: CÓDIGOS DEL ANÁLISIS DE LAS ENTREVISTAS	258

ÍNDICE DE TABLAS

Tabla 3.1. Etapas de investigación.....	72
Tabla 3.2. Rango de edad de participantes.....	77
Tabla 3.3. Género de los participantes.....	78
Tabla 3.4. Años de experiencia en docencia de lenguas de los participantes.....	79
Tabla 3.5. Escolaridad de los participantes.....	80
Tabla 3.6. Participantes por organismo académico de adscripción.....	82
Tabla 3.7. Participantes en las entrevistas.....	84
Tabla 3.8. Clasificación de reactivos de encuesta por categorías.....	86
Tabla 3.9. Valores para opciones de respuesta de cuestionario.....	92
Tabla 4.1. Porcentaje de respuestas por categoría inicial.....	104
Tabla 4.2. Estadística descriptiva para el cuestionario de prácticas evaluativas y diseño de exámenes.....	106
Tabla 4.3. Análisis factorial exploratorio: Varianza total explicada.....	109
Tabla 4.4. Factor 1: Valores e impacto.....	110
Tabla 4.5. Factor 2: Responsabilidad en el proceso de evaluación.....	111
Tabla 4.6. Factor 3: Aspectos técnicos.....	112
Tabla 4.7. Factor 4: Propósitos y objetivos del examen.	112
Tabla 4.8. Factor 5: Utilidad del examen.....	113
Tabla 4.9. Factor 6: Preparación del examen.....	113
Tabla 4.10. Análisis factorial exploratorio: Factores resultantes.....	114
Tabla 4.11. Tabla de co-ocurrencias de los diez códigos más con incidencias.....	118
Tabla 4.12. Tabla de incidencias de códigos por participante.....	126
Tabla 4.13. Tabla de incidencias de códigos por dimensión y participante.....	126

ÍNDICE DE FIGURAS

Figura 1.1. Niveles diferenciados de literacidad en evaluación de lenguas de acuerdo con el sector de involucramiento.....	27
Figura 1.2. Perfil de literacidad en evaluación del profesor y del diseñador de exámenes de lengua.....	28
Figura 3.1. Procedimiento de la fase cuantitativa de la investigación.....	90
Figura 3.2. Procedimiento de la fase cualitativa de la investigación.....	91
Figura 3.3. Ejemplo de codificación abierta.....	96
Figura 3.4. Reporte de códigos (extracto).....	97
Figura 3.5. Lista de vínculos código-código (Extracto).....	99
Figura 3.6. Red de Dimensiones y códigos relacionados.....	100
Figura 4.1. Relación entre los diez códigos más con incidencias.....	119
Figura 4.2. Dimensión lingüística.....	121
Figura 4.3. Dimensión técnica.....	122
Figura 4.4. Dimensión ética.....	123
Figura 4.5. Dimensión personal.....	124
Figura 4.6. Dimensión docente.....	125
Figura 4.7. Lista de vínculos código-código (Extracto).....	127
Figura 4.8. Red de Dimensiones y códigos relacionados (Vista extendida).....	128
Figura 4.9. Red de Dimensiones y códigos relacionados. (Vista circular).....	129

ÍNDICE DE GRÁFICAS

Gráfica 3.1. Rango de edad de participantes.....	78
Gráfica 3.2. Género de los participantes.....	79
Gráfica 3.3. Años de experiencia en docencia de lenguas de los participantes.....	80
Gráfica 3.4. Escolaridad de los participantes.....	81
Gráfica 3.5. Participantes por organismo académico de adscripción.....	83
Gráfica 5.1. Dispersión de datos de cuestionario (Elaboración propia).....	107

LISTA DE ACRÓNIMOS Y ABREVIATURAS

AFE - Análisis factorial exploratorio

CELe - Centro de Enseñanza de Lenguas

CENNI - Certificado Nacional de Nivel de Idioma

DAL - Dirección de Aprendizaje de Lenguas

DEL - Diseño de exámenes de lengua

ETS - Educational Testing Services

MCRE - Marco Común de Referencia Europeo

L2 - Segunda lengua

LEL - Literacidad en evaluación de lenguas

RP - Responsabilidad profesional

UAEMex - Universidad Autónoma del Estado de México

INTRODUCCIÓN

El origen de este trabajo se remonta al interés del investigador por la enseñanza de lenguas y principalmente por la evaluación del aprendizaje de éstas. La experiencia de elaboración de exámenes, tanto de manera individual, como en colaboración con colegas profesores y la asesoría de alumnos para la preparación de evaluaciones condujo a varios cuestionamientos sobre lo que se llamará la práctica evaluativa: ¿Es posible establecer que un instrumento dado es un buen examen? ¿Por qué hay aspectos de la evaluación en que hay coincidencias entre colegas y otros en los que no hay acuerdos? ¿Por qué algunos profesores diseñan intencionalmente reactivos poco claros a los alumnos? ¿Por qué algunos profesores no diseñan exámenes? ¿Cuál es la visión que tienen los profesores de los exámenes que ellos mismos diseñan? ¿Cuál es la visión que tienen los profesores de los exámenes elaborados por sus colegas?

Estos cuestionamientos sumados a literatura enfocada en el aspecto social del diseño de exámenes como Bachman y Palmer (1996, 2010), McNamara (2000), McNamara y Roever (2006), donde se hace un llamado a que los diseñadores de exámenes tomen responsabilidad por los instrumentos elaborados, aunque por supuesto lo hicieron con un enfoque en otros contextos diferentes al del salón de clases de segundas lenguas en México. En los textos mencionados la validez es una cualidad indispensable de un instrumento justo, este concepto y el de la responsabilidad social desarrollaron un interés por lo que había detrás de los exámenes que se aplican en el salón de clases y conocer con base en qué elementos o principios los profesores tomaban las decisiones referentes a los instrumentos aplicados a los alumnos. La literatura sobre el diseño de exámenes en el salón de clases era escasa en esos años, así como estudios específicos en contextos mexicanos. Lo anterior ha llevado al investigador a plantear como objetivo de este trabajo deconstruir la práctica evaluativa y de diseño de exámenes de lengua.

El contexto en que se llevó a cabo este estudio, una universidad pública en el Estado de México, fue elegido por conveniencia, dada la condición laboral del investigador, como

profesor en esta universidad y debido a la posibilidad de tener acceso a una cantidad de profesores de lengua suficiente para realizar el estudio.

La Universidad Autónoma del Estado de México (UAEMex) tiene presencia en 24 municipios del Estado de México. De acuerdo con la Agenda Estadística de la UAEMex (UAEMex, 2022), en el ciclo 2021-2022 su matrícula en estudios a nivel superior era de 59,665 estudiantes y 3,541 en estudios avanzados y un total de 7,715 profesores, incluyendo profesores de tiempo completo, medio tiempo y asignatura. La Universidad cuenta con 22 facultades, 17 centros universitarios, 22 centros de investigación, que albergan 84 programas de licenciatura, 42 especialidades, 37 programas de maestría y 22 de doctorado. La mayor parte de la población universitaria se concentra en el valle de Toluca, principalmente en tres campus: el Cerro de Coatepec, el área de Colón y el Cerrillo.

El objeto de estudio de este trabajo es el profesor de lengua extranjera en la Universidad Autónoma del Estado de México (UAEMex). La lengua inglesa es obligatoria para todos los planes de estudio de licenciatura, y la lengua francesa lo es en la Facultad de Lenguas. Con excepción de dicha facultad, cada plantel cuenta con un coordinador de lenguas, quien, a su vez, está coordinado por la Dirección de Aprendizaje de Lenguas (DAL), dependencia de la Secretaría de Docencia de la administración central de la UAEMex. En esta dirección se coordina el trabajo de los profesores de lengua y es donde se aprueban los planes de estudios. La UAEMex tiene como meta institucional el egreso de los alumnos de licenciatura con nivel B1 del Marco Común de Referencia Europeo (MCRE).

Fuera del ámbito de la DAL, también se encuentra el Centro de Enseñanza de Lenguas (CELe), que depende de la Secretaría de Extensión y Vinculación y que ofrece clases de lengua al público en general. La DAL reporta que en 19 facultades 174 profesores de lengua imparten clases, además de 34 profesores en la Facultad de Lenguas, para un total de 208 profesores de lenguas a nivel licenciatura en la UAEMex en el valle de Toluca.

La gran mayoría de los profesores de lengua inglesa en la UAEMex trabaja por horas clase, o como profesor de asignatura, como se le denomina en la institución. Algunos laboran en un solo plantel, aunque otros trabajan en dos o más y puede ser en educación superior o educación media superior, lo que ocasiona que un profesor tenga que desplazarse de un plantel a otro constantemente (Moore, Grondin y Andrade, 2020). El número de horas clase

de muchos de estos profesores llega a ser de 30 a la semana, el máximo que permite la Universidad y a esto hay que sumar las horas-clase que puedan impartir en instituciones privadas o de otro nivel educativo.

El perfil de los profesores de lengua en la Universidad es variado, los más experimentados, con más de veinte años impartiendo clases de lengua, son profesionistas de distintas áreas que por alguna razón decidieron tomar la carrera de enseñanza de lenguas. Otros perfiles son los de los Licenciados en Lenguas, que cuentan con un grado de licenciatura y que con el tiempo y los requerimientos como, justamente, tener carrera afín han podido colocarse cada vez con más frecuencia. No obstante, también hay entre los profesores de lengua nuevas contrataciones de profesionistas que cuentan con el nivel de lengua suficiente para obtener una plaza.

Cada facultad agrupa a sus profesores de lengua en áreas de docencia o academias, y cada una de estas a su vez cuenta con un presidente que dirige el trabajo colegiado, y parte de éste es coordinar el desarrollo de exámenes. Este estudio se enfoca en estos profesores de lengua que diseñan exámenes para sus alumnos de licenciatura. En cada plantel el coordinador organiza el trabajo en academia de todos los profesores. En estas academias, de manera independiente por plantel, se acuerda quienes diseñarán los instrumentos de evaluación por período. En algunas facultades se decidió que cada profesor es responsable del examen para su grupo.

La Legislación Universitaria (UAEMex, 1984) indica que cada semestre se deberá evaluar el aprendizaje del alumno en dos periodos parciales, un ordinario y, en caso de ser necesario, una evaluación extraordinaria o a título de suficiencia. Las academias de cada plantel deciden la manera en que se llevarán a cabo estas evaluaciones; en algunos casos, como la Facultad de Medicina, la manera en que se diseñan y aplican los instrumentos es dictada por las políticas del plantel o facultad.

A lo largo de este trabajo se hará referencia constantemente a “el profesor”, es preciso aclarar que cuando así sucede, se entiende por el profesor o profesora de lengua extranjera que labora en la universidad mencionada y principalmente en el papel de evaluador del aprendizaje, como diseñador de exámenes y otros instrumentos de evaluación,

exclusivamente para sus cursos y no como diseñador de reactivos para exámenes a gran escala.

Una vez establecido el contexto y el objeto de estudio, se planteará la problemática.

El objetivo general de este trabajo fue deconstruir la práctica evaluativa del profesor de lengua como diseñador de exámenes de lengua extranjera. Por práctica evaluativa se entiende todas las acciones relacionadas con la planeación, diseño, aplicación y calificación de exámenes u otros instrumentos de evaluación del aprendizaje de lenguas, así como el reporte de sus resultados. Conforme avanzó la investigación, se concluyó que era necesario ampliar el objetivo a conocer también la responsabilidad profesional del profesor como evaluador del aprendizaje de lengua.

La responsabilidad profesional es un concepto que surge como resultado de la metodología de la teoría fundamentada, aplicada en este estudio, y propuesto por el investigador para aglutinar los factores que se conjugan e influyen en las decisiones que toma el profesor de lengua con respecto a la práctica evaluativa.

La investigación se desarrolló en dos fases: cuantitativa y cualitativa. En la fase cuantitativa se encuestó a 88 profesores de lengua extranjera, inglés y francés, en la universidad. La encuesta se enfocó en conocer las prácticas evaluativas y de diseño de exámenes del aprendizaje de los profesores, a través de una escala de frecuencia de tipo Likert. La fase cualitativa consistió en entrevistas individuales semiestructuradas con doce profesores, quienes también participaron en la fase cuantitativa y accedieron de manera voluntaria a ser entrevistados. Como metodología de análisis se eligió la teoría fundamentada. Se transcribieron las entrevistas y se procedió al análisis del discurso de los profesores por medio del software Atlas TI.

El trabajo que se presenta consiste en dos capítulos que funcionan como marco de referencia teórico, el capítulo I sobre evaluación educativa, incluyendo evaluación de lenguas, evaluación de lenguas en el salón de clase y diseño de exámenes de lengua y literacidad en evaluación de lenguas. Esta última se entiende como las habilidades, conocimiento y principios que debe tener un individuo involucrado en un proceso de

evaluación de lengua (Taylor, 2009), concepto que se considera pertinente al tema, dado que contempla los aspectos necesarios para la práctica evaluativa.

En el capítulo II se presenta la teoría referente a la ética, ética en la educación, y la ética en la evaluación de lenguas. Ambos capítulos teóricos constan de un inciso que relaciona su respectivo tema con la complejidad (Morin, 1990), dado que la teoría de la complejidad es la base epistemológica del estudio. Se decidió escribir un capítulo sobre ética, dada la relevancia del aspecto ético en el carácter social de la evaluación (Kunnan, 2000; McNamara y Roeber, 2006) y su presencia en los factores de decisión identificados en la investigación después de haberse llevado a cabo el análisis cuantitativo y confirmado en la fase cualitativa.

Una vez enmarcada la teoría, se describe la metodología de investigación, contextualización, descripción de los participantes e instrumentos, así como de las fases del trabajo de investigación. Como enfoque metodológico se optó por el enfoque mixto secuencial exploratorio (Creswell y Plano Clark, 2018), que consiste en dos fases consecutivas de distinto corte, cualitativo y cuantitativo que se complementan entre sí. El estudio fue de diseño de tipo no experimental, transversal y descriptivo.

En la primera fase, cuantitativa, se diseñó un cuestionario que se aplicó a 88 profesores de lengua de la universidad, con el que se buscaba identificar y conocer las prácticas en evaluación de los profesores. En la segunda fase, con enfoque cualitativo, se entrevistó a 12 profesores, en una entrevista semiestructurada y se les preguntó acerca de sus prácticas en evaluación, particularmente en diseño de exámenes. Estas entrevistas se transcribieron y analizaron con base en el método cualitativo de la Teoría Fundamentada (Strauss y Corbin, 1998). De igual manera, se da cuenta de cómo se llevó a cabo la metodología en la tradición mencionada.

Como base epistemológica se decidió adoptar una visión desde la teoría de la complejidad, en la que debe observarse un fenómeno en la totalidad de sus componentes y factores. Morin (1990) explica que la ciencia ha fallado en explicar el conocimiento debido a su fragmentación en especialidades, que dejan de ver las relaciones entre ellas, entre diferentes factores que pueden intervenir en cómo se lleve a cabo dicho fenómeno. Esta epistemología también contempla el desarrollo de sistemas complejos como unidades que se autorregulan

y adaptan a los cambios en los factores que lo componen y en los factores externos. Estos sistemas interactúan unos con otros, sumando a la complejidad de su comportamiento. Como se explicará con detalle en su oportunidad, en este trabajo tanto la práctica evaluativa como la responsabilidad profesional se perciben como sistemas complejos.

Finalmente, se considera necesario un estudio sobre la práctica evaluativa ya que es necesario dar cuenta de las acciones realizadas por los profesores de lengua en el contexto mexicano. Algunos estudios describen prácticas en otros países (Vogt y Tsagari, 2014, Wang, 2017), pero la documentación de la práctica evaluativa del aprendizaje de lenguas en México es escasa. Adicionalmente, en el enfoque metodológico adoptado, se logra una introspección de los principales actores del proceso de evaluación, los profesores.

Con esto también se contribuye a lograr una percepción del carácter social de la evaluación que promueven Bachman y Palmer (1996) y McNamara y Roever (2006) y que debería repercutir en una práctica evaluativa más profesional, ética y socialmente responsable.

CAPÍTULO I

EVALUACIÓN EDUCATIVA Y DE LENGUAS

Introducción al capítulo

La lingüística aplicada es el área de conocimiento que busca contribuir a la solución de problemas relacionados con la lengua (James, 1993), como es el proceso de enseñanza-aprendizaje de lenguas. Una parte fundamental de este proceso educativo es la evaluación, ya que en las instituciones educativas es necesario tomar decisiones de diversa índole y a diferentes niveles jerárquicos y las evaluaciones proporcionan los datos o información que se requieren para tomar dichas decisiones. En general, los procesos de evaluación inician con la detección de una necesidad o establecimiento de un objetivo, se elabora o adapta un instrumento para obtener la información que se requiere, se aplica el instrumento, se analizan los datos y se toman decisiones.

Dada la relevancia de las decisiones que se toman en relación con este proceso, es imperativo para los especialistas en lingüística aplicada, especialmente aquellos dedicados a la enseñanza de lenguas y diseño de exámenes de lengua, adquirir responsabilidad sobre los instrumentos que se utilizan, para que las decisiones sean tomadas con base en instrumentos diseñados con un fundamento ético, un argumento y justificación válidos, y profundo conocimiento de la lengua a evaluar, esenciales en un docente o diseñador de exámenes, lo que denominamos en este trabajo *actuar con responsabilidad profesional*. Varios autores (Messick, 1989; Bachman y Palmer, 1996, 2010; Davies, 2008; Shohamy, 2000; McNamara y Roever, 2006; Chapelle, 2013) destacan la importancia de que el profesional de lengua actúe con estas bases. Este trabajo es una contribución a ese señalamiento; es una mirada de un docente de lengua e investigador acerca una problemática de la lingüística aplicada, donde es necesario precisar cómo se da esa responsabilidad en el profesional, quien toma decisiones en el momento de evaluar a sus estudiantes y diseñar los instrumentos para esta evaluación. Conocer el contexto en el que se desenvuelve el docente de lengua, iniciando por el proceso educativo en general, es esencial para un mejor entendimiento de la problemática presentada.

Es preciso advertir que, a lo largo de este capítulo, y de todo el estudio, se mencionarán tres conceptos fundamentales en evaluación que son validez, confiabilidad y practicidad, que son parte de las bases referidas renglones arriba para el profesional en la enseñanza y evaluación de lenguas. En páginas subsecuentes se tratarán estos conceptos con mayor profundidad, pero con el fin de proveer mayor claridad se tomarán las siguientes definiciones para: practicidad, que se refiere a los recursos necesarios para el diseño, aplicación y calificación de un examen (Brown y Abeywickrama, 2018, p.28); por confiabilidad se entiende la consistencia en la medición por un instrumento (Hughes, 2003, p.3); y por validez, la precisión con que un examen mide lo que se pretende medir (Hughes, 2003, p. 26).

En los siguientes párrafos de este trabajo se revisan y analizan las investigaciones con mayor impacto y las más recientes en cuanto a evaluación de lengua y diseño de exámenes. Se pretende que el lector encuentre en las siguientes líneas las bases para un mejor entendimiento del mundo del profesor de lengua, en especial aquel que tiene que diseñar o adaptar, por convicción u obligación, instrumentos que permitan obtener información de los estudiantes, en específico, acerca de su competencia comunicativa; y que sirva para el profesor para una reflexión sobre su propia práctica evaluativa.

1. Evaluación

Bachman y Palmer (2010) hacen la distinción entre los términos en inglés *assessment* y *evaluation*, refiriéndose al primero como el proceso de recopilar información sobre algo de nuestro interés, de una manera sistemática y fundamentada; y el segundo como una actividad que involucra elaborar juicios de valor y tomar decisiones con base en la información obtenida a través de la medición o *assessment* (Bachman y Palmer, 2010, p. 201-221). Mateo y Martínez (2008) utilizan el término ‘valoración’ para referirse a *assessment*. En este trabajo se utilizará el término evaluación para referirse al proceso de obtener información y utilizarla para tomar decisiones, agrupando los términos en inglés *assessment* y *evaluation*. Esta decisión no es arbitraria, en la práctica educativa es común hacer esta aglutinación, mientras que en el campo de la psicometría hacer la distinción es indispensable.

Hay que precisar que la evaluación educativa no se limita a la medición de los conocimientos o competencias adquiridas por los estudiantes, Tagliante (2005), por ejemplo, habla de tres funciones de la evaluación: pronosticar, diagnosticar e inventariar o certificar; para obtener la información que les permita tomar decisiones con respecto a dichas funciones los participantes o interesados en los procesos educativos se valen de la medición (*assessment*).

Cabra-Torres (2010) describe el discurso sobre la evaluación educativa a través de las últimas décadas y encuentra dos posturas, la primera implica que la evaluación es la emisión de un juicio de valor, que otorga un poder al evaluador para opinar sobre el evaluado. La segunda postura concibe la evaluación como un proceso donde se construye una descripción y una narrativa que proporciona información acerca del evaluado y que servirá de base para la toma de decisiones.

Mateo y Martínez (2008) abundan sobre la complejidad del proceso de evaluación educativa:

Bajo el término castellano 'evaluación educativa', coexiste un conjunto de prácticas y enfoques tan diversos que a su vez exigen, para su desarrollo, la aplicación de metodologías tan distintas y diferenciadas que han conducido a una división disciplinar basada en sus ámbitos de aplicación. Así hablamos de evaluación del aprendizaje de los estudiantes, de la docencia, de programas, de centros educativos, institucional y del sistema educativo entre otras... (Mateo y Martínez, 2008, p.27).

Por lo que, de acuerdo con estos autores, se requiere de especialistas que lleven a cabo la obtención de la información apropiada y oportuna, para una toma de decisiones informada y sustentada que coadyuve al pronóstico, diagnóstico o certificación de algún participante o proceso que conforman la práctica educativa.

Lo anterior no discrepa con el concepto de evaluación educativa propuesto por Pimienta (2008, p.4):

... concebimos la evaluación educativa como un proceso sistemático de recopilación de información (cualitativa y/o cuantitativa) para enjuiciar el valor o mérito de algún ámbito de la educación (aprendizajes, docencia, programas, instituciones, sistemas nacionales de educación), previa comparación con unas normas o criterios determinados con anterioridad y que responden a instancias de referencias específicas

Es necesario resaltar la importancia de la evaluación educativa; se habla de toma de decisiones, las cuales repercutirán, en distintos grados, en las vidas de los participantes en los procesos. Aunado a ello, habrá que considerar que a través de la práctica evaluativa se puede tener una visión más amplia o entender mejor los procesos educativos, con miras a su transformación, como lo expresa Cappelletti (2004); lo cual no es tarea fácil; el cambio del proceso educativo implica modificar el contexto, lo cual, a su vez, requiere de más información, con mejor sustento, voluntad de los actores y una fuerte reconsideración de los paradigmas educativos vigentes.

El proceso de evaluación educativa se lleva a cabo en todos los niveles, sin embargo, para este trabajo se delimitará el término evaluación o evaluación educativa a la medición y evaluación de los aprendizajes, conocimientos o competencias de los estudiantes de lengua. A continuación, se presentará una visión de la problemática que se encuentra al analizar la situación de este tipo de evaluación como resultado de los mencionados paradigmas.

1.1 Evaluación desde la perspectiva de la complejidad.

La evaluación educativa es un proceso complejo. La complejidad no es sinónimo de dificultad, sino que los factores, participantes, hechos, relaciones, consecuencias, o como menciona Celin Vargas (2012) quien habla de una combinación de habilidades, destrezas aptitudes y valores, además de aprendizajes metodologías, concepciones individuales y epistemologías opuestas se entrelazan de tal manera, que no es posible tener una visión de una problemática analizando de manera aislada uno de sus componentes (Morin, 1990). Son muchas las personas que intervienen en el proceso evaluativo, generalmente con perspectivas e intereses diferentes.

Un ejemplo de complejidad es el proceso evaluativo. Este es complejo desde un inicio, en el momento en que se tiene que definir qué es lo que se va a evaluar, definir el constructo a medir. Éste se manifiesta a través de una competencia o una habilidad definida *a priori* y, a su vez, en la evaluación de lenguas, es necesario obtener una muestra suficiente de esa manifestación, lo que le da validez de constructo a las evaluaciones. Esta manifestación debería representar el uso real de la lengua y adecuado al contexto del estudiante o del candidato, así que es necesario construir un argumento de validez (Kane, 2001) para la evaluación, aunque a la fecha no hay consenso sobre cómo desarrollar uno, a pesar de los

constantes esfuerzos de Messick (1989), Kane (2001), Bachman y Palmer (2010) y Chapelle (2011) entre otros.

En el ámbito educativo se buscan evaluaciones justas y válidas, pero a la vez tienen que ser prácticas, confiables y apegadas a los paradigmas educativos vigentes; así es como estos principios, que se contraponen, originan un problema complejo.

La practicidad de una evaluación se consigue, generalmente, a costa de la validez; y lograr confiabilidad representa un alto costo, lo que se contrapone a la practicidad, una evaluación auténtica puede tener un grado de validez mucho mayor que una prueba de opción múltiple, pero a su vez, es poco práctica. Sumemos a esto las políticas y posturas institucionales, las perspectivas y paradigmas individuales y necesidades de los involucrados, tenemos como resultado una complejidad, que es un denominador común a un infinito de contextos y situaciones. Las propuestas de solución generales no tienen el alcance necesario, son insuficientes, excluyentes y parciales; cuando una evaluación puede ser de la mayor importancia para cada uno de los participantes, la generalidad no les cumple.

1.2 Evaluación de aprendizaje de lenguas.

El proceso de enseñanza-aprendizaje de lenguas, como parte del proceso educativo, tiene la misma problemática, la evaluación de lenguas se lleva a cabo bajo los mismos paradigmas que el resto de la educación. La evaluación tradicional prevalece ante las propuestas constructivistas, así que las instituciones, los profesores y el resto de las personas interesadas, (padres de familia, comités evaluadores, etc.) buscan y pretenden medir, o que se mida, el conocimiento de la lengua por medio de exámenes, como si fuera la única alternativa.

1.2.1 Enfoques en evaluación de lenguas.

La evaluación del aprendizaje de lenguas se ha tratado desde diversas perspectivas o enfoques, debido a los usos que se les da a los resultados: enfoque general, enfoque en la enseñanza, enfoque social y crítico de la evaluación y en la acreditación y certificación, los cuales se comentan a continuación.

Varios autores han provisto una visión detallada y muy precisa sobre la evaluación de lenguas (Davidson, 2012; Green, 2013) y presentan, tanto para maestros de lengua como diseñadores de exámenes de certificación, una perspectiva sobre la evaluación y el diseño de exámenes de lengua. Green (2013) explica cómo evaluar las habilidades productivas y receptivas y ayuda a la comprensión del constructo de competencia lingüística. También en el área de evaluación de lenguas, existen una serie de compilaciones de artículos (Shohamy, 2000, Kunnan, 2013) de expertos en temas diversos y especializados temas como validez, procesos de evaluación, diseño de exámenes, evaluación de habilidades, marcos de referencia, ética y evaluación, entre otros.

Los autores de metodología de enseñanza (Ur, 2012; Harmer, 2015) lo estudian, de manera breve, como una parte más del proceso enseñanza-aprendizaje de lengua. Estos autores proponen metodologías y tareas para evaluar las habilidades productivas y receptivas. En el mismo sentido, pero con mayor profundidad Brown (2003) y Brown y Abeywickrama (2018) exploran la evaluación desde un enfoque del docente de lenguas en el salón de clase. Estos autores sientan bases teóricas para cada habilidad, proponiendo subhabilidades, tipos de producción o desempeño y una serie de tareas que el docente puede utilizar para evaluar a sus alumnos. Si bien el enfoque de Brown es en la evaluación en general (*assessment*), es notable el peso que tienen los exámenes de lengua en su metodología de evaluación.

En la serie *Assessing*, editada por Charles Alderson, se explora a profundidad la evaluación de las habilidades, producción escrita (Cushing Weigle, 2002), producción oral (Luoma, 2004) comprensión escrita (Alderson, 2000) y comprensión oral (Buck, 2001), además de la evaluación para niños (McCay, 2006), gramática (Purpura, 2004) vocabulario (Read, 2000) para propósitos específicos (Douglas, 2000) y lengua a través del uso de tecnología (Chapelle y Douglas, 2006). Esta serie constituye un referente importante en la concepción de instrumentos de evaluación, tanto para docentes, como para diseñadores de exámenes en gran escala o de certificación y el enfoque de la serie, en el constructo a evaluar, ha permitido desarrollar el estudio crítico del impacto social de la evaluación de lenguas y su validez, que es la cualidad de un examen de medir para lo que fue diseñado (Bachman y Palmer, 1996). El enfoque crítico y social en evaluación de lenguas y en diseño de exámenes es detonado, precisamente, por este concepto.

Messick (1989) establece la importancia de la validez en los exámenes de lengua y hace un llamado a reconocer la importancia del examen en las consecuencias que tiene el resultado para el candidato. Bachman (1990) y después Bachman y Palmer (1996, 2010) y Bachman y Dämbock (2018) proponen la elaboración de un argumento de uso de un examen de lengua, aunque estos últimos se dirigen al profesor de lengua y los exámenes y evaluaciones para la clase de segunda lengua (L2), principalmente.

Por otro lado, el enfoque crítico y social de la evaluación de lenguas pugna porque los evaluadores realicen los instrumentos con responsabilidad, dada la relevancia de los resultados de dichas evaluaciones. Es necesario que el evaluador y diseñador de exámenes tenga una conciencia y actúe considerando el aspecto social de la evaluación (Shohamy, 1997; McNamara, 2000; Shohamy, 2001; McNamara y Roever, 2006). Este enfoque tiene como principal objeto de crítica los exámenes de lengua, cuyas características y procesos de elaboración se abordarán a detalle en el siguiente apartado de este capítulo; y es que los exámenes de lengua no solamente son cuestionables por el proceso de elaboración, sino por el uso que se les da. Los exámenes se han vuelto instrumentos de poder, que las instituciones aplican mediante organismos acreditadores y no siempre con la rigurosidad que se requiere dada la importancia y que además excluyen, etiquetan y condicionan a los individuos conforme a las políticas de quienes se encuentran en posiciones de poder (Shohamy, 2007).

El enfoque crítico (Shohamy, 2001; McNamara y Roever, 2006) también sostiene que debido a la importancia de muchas de las decisiones que se toman con base en estos exámenes y la complejidad que conlleva la elaboración de un instrumento de evaluación de lengua, es deseable que los instrumentos tengan un alto grado de validez y promover la cultura de la examinación. En otras palabras, con frecuencia los actores a cargo de esas decisiones no tienen el conocimiento sobre estas dificultades; en la generalidad se piensa que cualquier instrumento para evaluar lengua sirve para cualquier propósito, sin embargo, cada instrumento debería medir la competencia comunicativa con una muestra significativa, una definición clara del constructo y de uso de la lengua, conforme a la necesidad del usuario y de los interesados en los resultados de las evaluaciones. Es así, que, la validez se vuelve un eje central y más allá de la validez justificada psicométrica y estadísticamente, el enfoque crítico se ocupa del ser humano, de sus necesidades sociales, busca promover en el

evaluador una conciencia ética y de empatía con la sociedad, de búsqueda de equidad y justicia a través de instrumentos más válidos, confiables y por consecuencia más éticos.

Otro enfoque de la evaluación es en la certificación de idioma, que se ha vuelto una necesidad en diversos ámbitos: a los docentes les solicitan certificarse como prueba de su competencia comunicativa, por ejemplo el Certificado Nacional de Nivel de Idioma (CENNI); en las universidades es necesario demostrar que se cuenta con los medios lingüísticos para que un estudiante sea admitido en los programas de estudios (Alderson y Hamp-Lyons, 1996; Green, 2006; Deygers, Van den Branden y Van Gorp, 2017); para los migrantes es indispensable probar el dominio de la lengua oficial del país donde pretenden obtener una nacionalidad (Krumm, 2007); y los estudiantes de lengua se ponen como meta obtener una certificación como muestra de un logro. Estos escenarios, aunque diferentes, requieren todos de un alto grado de confiabilidad en los instrumentos, y las instituciones a cargo de certificar la lengua, como Cambridge Assessment o Educational Testing Services (ETS), buscan precisamente eso, alto grado de confiabilidad mediante un estricto control de las aplicaciones de los exámenes de certificación, en centros autorizados para ello.

Este enfoque ha generado una industria millonaria a nivel mundial, que además tiene un impacto directo en los salones de clases, ya que en diversos contextos la preparación de alumnos para exámenes de certificación se ha adoptado como una práctica regular, agregando un elemento más en la ya compleja práctica evaluativa, en la que el docente debe incorporar a su competencia el conocimiento de estos instrumentos de evaluación.

Otras instituciones, como universidades y editoriales, también buscan desarrollar sus propios instrumentos de certificación, con la limitante presupuestal, ya que el desarrollo y aplicación de instrumentos de evaluación de lengua requiere de una inversión importante de recursos. Las escuelas de idiomas, así como las instituciones educativas, necesitan implementar procesos de desarrollo de exámenes, ya sea para certificación del idioma o acreditación de la competencia lograda después de un curso, y quienes generalmente están a cargo de estos procesos son profesores de lengua.

Debido a esto, surge la necesidad de desarrollar la cultura de la literacidad en evaluación y fomentar la actuación evaluativa en un marco de responsabilidad profesional. Como se puede ver en estas líneas, los distintos enfoques mencionados convergen en esta necesidad

y aunque el paradigma constructivista promueve una evaluación más auténtica y de carácter más formativo, es la evaluación sumativa, a través de exámenes de tipo cuantitativo la que prevalece, debido a los innumerables factores que tienen efecto en la práctica del docente y en el proceso de evaluación del aprendizaje que le llevan a tomar decisiones variables, dependiendo del contexto particular. En la siguiente sección de este trabajo se presenta la teoría y literatura acerca de evaluación de lengua en el salón de clase y diseño de exámenes de lengua.

1.2.2 Evaluación de lengua en el salón de clase

Por muchos años la evaluación en el salón de clase fue vista como una derivación de la evaluación de lengua con exámenes a gran escala y que los principios de ésta debían ser aplicados a la primera (Turner, 2017). Hughes (2001), por ejemplo, desarrolla con profundidad el diseño de exámenes de lengua, sin embargo, su metodología está basada, de igual manera, en evaluación a gran escala.

Por otro lado, enfoques de evaluación para el profesor en el aula, serían los desarrollados por Genesee y Upshur (1996) y Brown y Abeywickrama (2018), quienes presentan procesos de diseño de instrumentos de evaluación de manera más práctica y cercana al profesor. A pesar de la tendencia a obviar la evaluación en el aula como si fuera lo mismo que la evaluación a gran escala, en los últimos veinte años la investigación sobre la primera ha tenido un importante desarrollo buscando hacer notar la diferencia entre ambos enfoques (Hill, 2017).

La evaluación de lenguas en el salón de clase (*Classroom-based assessment*) es definida por Hill y McNamara como:

Cualquier reflexión del profesor (y/o de los alumnos) acerca de las cualidades del trabajo de un alumno (o grupo de alumnos) y el uso de esa información por parte del profesor con propósitos de enseñanza, aprendizaje, retroalimentación, reporte, administración o socialización” (Hill y McNamara, 2012).

De manera similar, Genesee y Upshur (1996) consideran que los aspectos del proceso de evaluación en el salón de clase son determinar el propósito de la evaluación, recolectar e interpretar información y tomar decisiones con base en ella, que son exigencias para el profesor de lenguas. Con esto en consideración, Genesee y Upshur (1996) presentan un

tratado amplio y práctico enfocado en la enseñanza de segundas lenguas, el profesor de lengua y el proceso de evaluación en el aula, con el que buscan ayudar al profesor en planear evaluaciones relevantes y efectivas; tomar decisiones acertadas y en volverse agentes de cambio en su salón de clase y en los programas de estudios.

Brown y Abeywickrama (2018) tratan el tema de la evaluación en el salón de clase, desde terminología básica hasta descripción de constructos, presentan el proceso de diseño, micro y macrohabilidades a evaluar, así como un amplio catálogo de tareas sugeridas para distintas etapas del aprendizaje de los alumnos.

Bachman y Damböck (2018) proponen un enfoque de evaluación de lenguas en el salón de clase en la misma línea de los trabajos previos de Bachman (Bachman y Palmer, 1996 y 2010), que busca proveer de mayor equidad y justicia en la evaluación, creando consciencia en el profesor de la responsabilidad y obligación de rendir cuentas sobre los instrumentos y evaluaciones que utiliza en el salón de clase. Esto se logra mediante la justificación de la evaluación que definen como el proceso en que los profesores como diseñadores y usuarios de exámenes tendrán que demostrar el grado en que la intención al usar un examen es justificada (Bachman y Damböck, 2018). Los autores proporcionan una variedad de ejemplos de este enfoque para la evaluación de distintas habilidades y en distintos niveles de competencia.

La literatura mencionada provee al profesor de una base teórica en cuanto a evaluación del aprendizaje en el salón de clases de lengua, no obstante, es necesario considerar que a pesar de la posible contribución que estas propuestas teóricas pudiesen aportar a su práctica profesional, cada contexto requiere de adecuaciones y adaptaciones que no son posibles de generalizar.

1.2.2.1 Evaluación formativa y sumativa

De acuerdo con Brown y Abeywickrama (2010:7) el profesor está evaluando constantemente, de manera formal e informal y coinciden con Turner (2017) en que la formal es relacionada principalmente con la evaluación sumativa y la informal con la formativa. La evaluación sumativa es aquella que se realiza al término de un periodo y busca medir el aprendizaje del alumno acerca de los contenidos cubiertos en dicho periodo.

La evaluación formativa está más ligada al proceso enseñanza-aprendizaje, en el que el profesor observa el desarrollo de las habilidades y competencias para contribuir en él y donde la retroalimentación del profesor se enfoque en logros futuros. Sobre el tema Rea-Dickins (2001) encontró que los profesores pasan más tiempo en evaluación informal, relacionada con la formativa que en evaluación formal, relacionada con la sumativa, y muchas decisiones que se toman en el aula dependen de esta evaluación.

Davison y Leung (2009) explican que los conceptos de evaluación basada en el salón de clase (*classroom-based assessment*), evaluación basada en el profesor (*teacher-based assessment*) y evaluación para el aprendizaje (*assessment for learning*) han sido utilizados por distintos autores para referirse al proceso de evaluación con fines pedagógicos, que es utilizada por el profesor para retroalimentar al estudiante y fomentar el aprendizaje. Los autores agregan que este tipo de evaluación debe ser desarrollada de principio a fin por el profesor, desde establecer e implementar los constructos, objetivos y tareas; se lleva a cabo en el salón de clases y por parte del profesor y, a diferencia de la evaluación sumativa, permite que a través de una variedad de tareas y actividades se pueda recolectar información a lo largo de cierto tiempo, además de que puede ser adaptada a las necesidades de los estudiantes. Wang (2107) utiliza el término *classroom assessment* para describir aquella que se lleva a cabo en el salón de clase y que incluye ambos tipos de evaluación, formativa y sumativa.

Brown y Abeywickama (2018) explican que la evaluación es parte de la enseñanza. Los profesores establecen las condiciones para que el alumno ponga a prueba sus hipótesis y conclusiones sobre la lengua, que intente producir, comprender, equivocarse, ser retroalimentado y volver a intentar, todo de manera informal, hasta que está listo para la evaluación formal, donde la competencia adquirida en el proceso tiene que ser demostrada por medio de las tareas del examen. Mientras tanto, el profesor observa, califica, provee retroalimentación, emite conclusiones, compara, discrimina, toma decisiones y elabora estrategias para que el alumno logre la competencia. De acuerdo con esto, el profesor evalúa permanentemente, durante la clase, en el día a día, asiente, congratula, corrige, motiva, de manera informal, o formativa, la participación del alumno, a través de las tareas pedagógicas en el salón de clase. Este proceso da lugar a la evaluación formal o sumativa. Brown y Abeywickama (2018) la definen como técnicas planeadas de muestreo

sistemático, diseñadas para proveer al profesor y al alumno una apreciación del desempeño de este.

Todas estas acciones relacionadas con la evaluación en sus distintas facetas ayudan a describir esta parte del trabajo del profesor de lengua, donde resaltan el carácter permanente de la evaluación formativa y la relevancia de ambos tipos de evaluación en la toma de decisiones de parte del profesor. Es necesario tener claro que la evaluación formativa y sumativa no se contraponen o que tampoco es posible declarar que una es mejor o más importante que la otra. Ambas son componentes de un mismo sistema del que se vale el docente para recopilar información, emitir juicios y actuar en consecuencia.

1.2.2.2 Evaluación y el enfoque comunicativo

En general, el constructo a medir en un examen de lengua es la competencia comunicativa (Bachman y Palmer, 1996). Brown y Abeywickrama (2018) explican que, con la llegada del método comunicativo para la enseñanza de lenguas surge también la necesidad de desarrollar tareas para evaluar las habilidades integradas y con mayor grado de autenticidad (que se define con detalle en un apartado más adelante en este capítulo, pero puede adelantarse como el grado en que las tareas de una evaluación asemejan a usos reales de la lengua), lo que implica más tiempo y recursos, para reemplazar las tareas artificiales que eran incapaces de proveer una muestra del desempeño del alumno en una situación que requiriera un uso real de la lengua. Las tareas comunicativas y más auténticas implican, a pesar de su costo, un grado de validez mayor.

Lo anterior, dio lugar a la evaluación basada en desempeño y a la evaluación basada en tareas. La primera consiste típicamente de tareas de producción oral y escrita, preguntas de respuestas abiertas, tareas grupales y de habilidades integradas. La segunda consiste en tareas interactivas que provocan que el alumno realice la función que se pretende desarrollar en él y en la que tiene que utilizar habilidades integradas para lograrlo (Brown y Abeywickrama, 2018). La evaluación se ha enfocado en los últimos años en el desarrollo de tareas más adecuadas a las características personales del alumno, a la evaluación con la ayuda de la tecnología y de técnicas de evaluación alternativa como los portafolios, presentaciones, entre otras, que permiten obtener un mayor grado de validez de los instrumentos y contribuyen a la evaluación para el aprendizaje.

El carácter comunicativo de la lengua es una constante que requiere que el profesor tenga presente en todo momento en su práctica. No obstante, existen limitantes, generalmente relacionadas con disponibilidad de recursos, o en otras palabras con la practicidad de un instrumento, que repercuten en el proceso evaluativo. La generación de tareas auténticas, más comunicativas depende de varios factores relacionados con este principio de evaluación, lo que contribuye a las numerosas variables en la práctica evaluativa. Ésta se describe en el inciso siguiente, desde una perspectiva teórica y con base en algunos estudios.

1.2.2.3 Prácticas evaluativas en el salón de clase

En este apartado se presenta la revisión de literatura, así como algunos antecedentes y conclusiones obtenidas de trabajos de investigación relacionados con la práctica evaluativa de los profesores de lengua en el salón de clase, que ha sido escasamente documentada ya que la mayoría de la investigación se ha enfocado en exámenes a gran escala que presentan una problemática distinta a la que enfrenta el profesor en el salón de clase (Turner, 2017). Sin embargo, temas relacionados con el proceso de evaluación (Brown y Abeywickama, 2018; Rea-Dickins, 2001); identidad (Rea-Dickins, 2001); y necesidades de formación (Hill, 2017).

Brown y Abeywickama (2018) presentan un proceso de diseño más adecuado a las necesidades del profesor en el salón de clases. Este proceso consta de las siguientes etapas:

1. Determinación del propósito del examen
2. Diseño de objetivo claros y específicos
3. Diseño de especificaciones
4. Diseño de reactivos
5. Aplicación del examen
6. Puntuación, calificación y retroalimentación

Es necesario precisar que por especificaciones, Brown y Abeywickrama, se refieren a un bosquejo de la estructura del examen o la evaluación, un documento que describa los objetivos, el contenido, el tipo de tareas y de reactivos, las habilidades a evaluar, cómo se calificará y cómo se reportarán los resultados a los alumnos. En cuanto al diseño de reactivos, Brown y Abeywickrama (2018) profundizan en la elaboración de reactivos de opción múltiple, advirtiendo de las debilidades de este tipo de reactivo. Los autores también sugieren un análisis de índices para analizar los reactivos, como el índice de facilidad/dificultad, índice de discriminación de reactivo y el índice de eficiencia del distractor.

Entre la literatura que reporta la investigación sobre la evaluación en el salón de clase de lengua destacan los trabajos recientes sobre evaluación para el aprendizaje. Davison y Leung (2009) revisan la literatura hasta ese momento conciliando los diferentes conceptos teóricos relacionados a este tipo de evaluación. Tan y Turner (2105) documentan el caso del uso de exámenes de producción escrita diseñados para aplicación en gran escala, pero que son aplicados y calificados por profesores a quienes se les pide hacer comentarios sobre el proceso de aplicación, las tareas y las rúbricas, lo que ayuda a los diseñadores a generar mejor validez de constructo y a los profesores involucrarse en el proceso y a llevar a cabo su instrucción en el salón de clase enfocada en el aprendizaje y en la producción requerida en la prueba. Wang (2017) documenta la evaluación de la producción oral realizada por una experimentada profesora de inglés como lengua extranjera en China. El estudio reveló que la profesora realizaba actividades de evaluación formativa que no eran vistas por ella de esa manera, sin embargo, para los alumnos contribuían de manera importante en la preparación para la evaluación sumativa.

Rea-Dickins (2001) entrevistó a tres profesores lengua inglesa como L2, de diferentes escuelas y se les solicitó describir los procesos de evaluación en el salón de clase. Este estudio contempla tanto evaluación formativa como sumativa. Las prácticas evaluativas se analizaron y describieron con base en los objetivos de la evaluación, las características de la actividad, el contenido o enfoque, conocimientos previos, operación, recursos y condiciones de evaluación. En los hallazgos se describe que los tres profesores comparten el mismo enfoque en la selección del contenido de la evaluación, que es en puntos muy específicos que se enseñaron de manera explícita. Se describen las estrategias de

implementación de la evaluación, dos de los tres ejemplos de evaluación fueron evaluación formativa e informal. De acuerdo con Rea-Dickins (2001) el ciclo de evaluación tiene cuatro etapas principales:

- I. Planeación
- II. Implementación
- III. Monitoreo y registro
- IV. Difusión de resultados.

Rea-Dickins (2001) presenta entre sus hallazgos diferentes identidades de la evaluación de lengua en el salón de clase:

- Identidad burocrática: relacionada con factores externos al aula que obligan al profesor a llevar a cabo las evaluaciones.
- Identidad pedagógica: Rea-Dickins plantea que esta identidad es interna y se relaciona con la necesidad del profesor de conocer sobre el desempeño de sus alumnos con fines de planeación de la clase.
- Identidad de aprendizaje: se relaciona con el aprendizaje del alumno y pretende involucrar al alumno, motivándolo y comprometiéndolo a través de una reflexión.

Hill (2017) concluye que los profesores no tienen claridad acerca de sus necesidades de preparación en cuanto a evaluación, lo cual se debe a una carencia de habilidades para reflexionar sobre la naturaleza y calidad de su propia práctica. Con base en esto, Hill y McNamara (2012) y Hill (2107) desarrollaron un modelo que permitiría al profesor reflexionar sobre su práctica y el cual se comentará más adelante.

En esta sección se han presentado los primeros intentos de conocer y explicar cómo es en la práctica el trabajo del profesor como evaluador, así como una breve revisión de la literatura dirigida específicamente a este. En este sentido, cobra importancia la literacidad en evaluación de lenguas, ya que ubicar al profesor en términos de su literacidad nos acerca a un entendimiento de su trabajo. Enseguida se discute ampliamente esta literacidad; se inicia definiendo la literacidad en evaluación en general, para continuar con la literacidad en

evaluación de lenguas; se comentarán algunos estudios trascendentes y se expondrá el alcance de este concepto en el ámbito particular de la evaluación de lenguas.

1.2.3 Literacidad en evaluación

En el ámbito de la educación en general, Stiggins (1991, 1995) propuso el término literacidad en evaluación para referirse al conocimiento y habilidades que se requiere en los profesores y administradores educativos en cuanto al área de evaluación. Aunado a esto, Popham (2009) enfatizó el aspecto ético en la evaluación al incluir en la literacidad en evaluación aspectos como confiabilidad, validez de contenido, equidad, evaluación auténtica, impacto de los resultados y la importancia de los exámenes y cómo afectan sus consecuencias al estudiante, y DeLuca (2012) propuso un modelo marco de referencia para la formación de profesores en el área de evaluación. De esta manera se ha desarrollado la conciencia de la profesionalización de sus participantes en el campo de la evaluación educativa.

1.2.3.1 Literacidad en evaluación de lenguas

De forma paralela, en la enseñanza de lenguas, la literacidad en evaluación ha generado reflexiones que nos permiten tener un marco de referencia. Dada la complejidad de un concepto como la lengua, su enseñanza tiene importantes particularidades, así que se requiere una serie de habilidades y conocimientos especializados (Davies, 2008) y como consecuencia, su evaluación también necesita que quienes estén a cargo de llevarla a cabo, posean un alto grado de especialización, es decir, literacidad en evaluación (*assessment literacy*). Ésta, de acuerdo con Davies (2008) y Taylor (2009), tiene tres componentes: habilidades metodológicas para el diseño de evaluaciones, conocimiento de la lengua y principios que regulen la conducta del evaluador.

La cultura de la literacidad en evaluación de lengua ha tenido un fuerte desarrollo en los últimos años, sin embargo, el camino ha sido largo. En el presente todavía no existe un consenso sobre lo que significa literacidad en evaluación de lengua. Se han realizado varios estudios sobre las necesidades de capacitación de profesores de lengua en el tema de evaluación (Davies, 2008; Fulcher, 2012; Grabowski y Dakin, 2013; Vogt y Tsagari, 2014; Lam, 2015; Inbar-Lourie, 2016, Giraldo, 2017, Vogt, Tsagari y Spanoudis, 2020) y de ellos

podemos concluir que, primero, los evaluadores de lenguas, haciendo una generalización, requieren de más preparación en el tema, tanto de diseño de exámenes como de instrumentos alternativos; segundo, la evaluación ha pasado la etapa donde la psicometría y el análisis estadístico predominaban en la “especialización” del evaluador, para llegar a un momento donde es necesaria la reflexión en temas éticos y humanísticos de la evaluación, como el uso adecuado de los exámenes, la consideración del contexto y características particulares y de los candidatos y la mejor preparación y conocimiento de la competencia lingüística por parte del evaluador o del diseñador de exámenes (Davies, 2013).

En este sentido, se debe resaltar la necesidad de que el evaluador actúe con responsabilidad (Shohamy, 1997; Hamp-Lyons, 2000; McNamara y Roever, 2006; Spolsky, 2008) ya que debido a la importancia de las decisiones que se toman con el resultado de los exámenes de lengua, incluso aquellos que no son exámenes de certificación sino evaluaciones sumativas de cursos de lengua en diversos niveles educativos.

Al hacer una revisión de los libros de texto sobre evaluación de lenguas, Davies (2008) encuentra que el enfoque de éstos había particularizado en dos aspectos: el conocimiento de la lengua, primero su descripción y cómo medirla, y segundo, las habilidades o técnicas de diseño y análisis de instrumentos de evaluación. Así que, al conocimiento y habilidades que debe tener el profesional de diseño de evaluaciones de lengua, Davies (2008) agrega lo que él denomina principios, que implican el uso apropiado de los exámenes, la equidad para todos los candidatos, impacto de los instrumentos y los aspectos éticos y profesionales. De aquí es donde Taylor (2009) define la literacidad en evaluación de lengua como los conocimientos, habilidades y principios que deben tener los profesionales de evaluación de lengua para ejercer.

A partir de Davies (2008), se han realizado varios estudios sobre la literacidad en evaluación de lenguas, con diferentes enfoques: Fulcher (2012), Scarino (2013) y Giraldo (2017) en el aula de enseñanza de lengua; Taylor (2009 y 2013) en los involucrados en el proceso de evaluación (*stakeholders*); el mismo Davies (2008) e Inbar-Lourie, (2008 y 2016) en la profesionalización del evaluador y desarrollo de sus competencias. A pesar de estos estudios, y otros más (Jeong, 2013; O’Loughlin, 2013; Pill y Harding, 2013; Vogt y Tsigari, 2014) no ha sido posible consensar un concepto o un modelo de análisis.

No obstante, la base establecida por Davies (2008) con respecto a lo que debe tener el evaluador de lenguas que son conocimiento, habilidades y principios, ha sido un eje sobre el cual diversos autores han basado su trabajo, con ciertas variantes.

Inbar-Lourie (2008:389) define literacidad en evaluación:

Tener literacidad en evaluación entonces, significa tener la capacidad de hacer y responder preguntas críticas sobre el propósito de una evaluación, acerca de la adecuación del instrumento a utilizarse, sobre las condiciones de examinación y sobre qué pasará con base en los resultados¹.

En esta definición, que se refiere a literacidad en evaluación en general Inbar-Lourie plantea cuatro aspectos a considerar: el propósito de la evaluación, la validez del instrumento, la confiabilidad y el impacto de los resultados. Inbar-Lourie agrega que al considerar las particularidades de la evaluación de lenguas y las competencias específicas de este campo, se vuelve necesaria la conceptualización de la literacidad en evaluación de lenguas, que la misma autora (Inbar-Lourie, 2016) y Taylor (2009:) definirían de manera más simplificada, como el conocimiento que requieren los involucrados en el proceso de evaluación para llevar a cabo evaluaciones de lengua. Pill y Harding (2013) la definen como la gama de competencias que habilitan a un individuo a entender, evaluar, diseñar y analizar datos obtenidos por exámenes de lengua. El concepto de O'Loughlin (2013, p.363) es más detallado:

La adquisición de una gama de habilidades relacionadas a la producción de un examen, la interpretación de los resultados y su uso, y la evaluación del examen, en conjunto con el desarrollo de un entendimiento crítico sobre los papeles y funciones de evaluación en la educación y la sociedad².

Este concepto de O'Loughlin muestra un enfoque importante en el aspecto social de la evaluación de lenguas y enfatiza dos aspectos, la producción o desarrollo de un examen y la evaluación del propio examen, sugiriendo que es necesario reflexionar sobre el uso que se le dé a los resultados del test y el impacto en el contexto social.

De igual manera, Fulcher (2012) incorpora el aspecto social y los contextos particulares a su propio concepto y modelo de literacidad en evaluación de lenguas. Este concepto

¹ Traducción del autor.

² Traducción del autor.

también tiene su origen en Davies (2008), sin embargo, es más específico (Fulcher, 2012:125):

The knowledge, skills and abilities required to design, develop, maintain, or evaluate, large-scale standardized and/or classroom-based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice. The ability to place knowledge, skills, processes, principles, and concepts within wider historical, social, political, and philosophical frameworks in order understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals.

Fulcher habla en primer lugar de conocimiento, habilidades y principios; es incluyente de ambas perspectivas, tanto de profesores como de examinadores profesionales; y en segunda instancia, menciona que es necesario ubicar estas características en un contexto específico, que considere marcos de referencia que permitan explicar las prácticas y su impacto en la sociedad.

El concepto de Fulcher es un principio en cuanto a ubicar el diseño de exámenes y evaluaciones en un proceso complejo se refiere, ya que consiste en una serie de componentes que son difíciles de entender de manera aislada o fuera de su contexto. Fulcher considera tres componentes principales: prácticas de diseño de exámenes, los principios que fungen como guía para la práctica y los contextos que ayudan a explicar orígenes, razones e impacto.

Giraldo (2017) propone otro modelo de análisis para la literacidad en evaluación de lengua. De igual manera se basa en tres componentes como plantea Davies (2008), conocimiento, habilidades y principios. El conocimiento se refiere a tres dimensiones: conocimiento sobre lingüística aplicada, sobre teorías y conceptos de diseño de exámenes y conocimiento del propio contexto. Por habilidades se refiere a la habilidad instruccional, en diseño de evaluaciones, en evaluación educativa y habilidades tecnológicas. Por principios se refiere a la conciencia a temas sensibles en evaluación de lenguas.

Para Hill (Hill y McNamara, 2012; Hill 2017) es indispensable que el profesor tenga plena conciencia sobre sus propias prácticas para poder identificar sus necesidades de preparación en evaluación. De acuerdo con la autora (Hill, 2017), los profesores no tienen claridad acerca de sus necesidades de preparación en cuanto a evaluación, lo cual se debe a una carencia de habilidades para reflexionar sobre la naturaleza y calidad de su propia práctica.

Hill y McNamara (2012) desarrollaron un marco conceptual como guía para los profesores para identificar y analizar su práctica evaluativa en el salón de clase. El modelo se basa en cuatro preguntas eje y una quinta que agregaría Hill más adelante (Hill, 2017):

- 1- ¿Qué hacen los profesores? (¿Qué prácticas?)
- 2- ¿Qué buscan con ellas?
- 3- ¿Qué teorías o estándares utilizan como base?
- 4- ¿Qué entienden los aprendientes sobre evaluación?
- 5- ¿Cómo influye el contexto en la práctica evaluativa?

El modelo de Hill hace una clasificación de las preguntas, señalando que las primeras dos se refieren a las prácticas evaluativa y pedagógica, la tercera a los conceptos sobre evaluación, y la cuarta y quinta al contexto. A diferencia de Giraldo (2017), el modelo de Hill no contempla los principios éticos que deben considerarse en la práctica.

Los modelos de literacidad de Fulcher (2012) y Giraldo (2017) presentan los conceptos más complejos, con un mayor número de componentes, incluyendo el contexto, el cual varía en cada individuo. Se considera que dichos modelos permiten un entendimiento más profundo de la práctica evaluativa y podrían convertirse en un punto de partida para una ligación más estrecha entre la práctica evaluativa y el proceso de enseñanza-aprendizaje. Los estudios y modelos mencionados en este inciso no aclaran el grado de especialización en evaluación que se espera del profesor de lengua. Enseguida, se presenta una revisión a la literatura que permitiría una visión más completa de las expectativas en el docente en este sentido y delinear parte del perfil profesional.

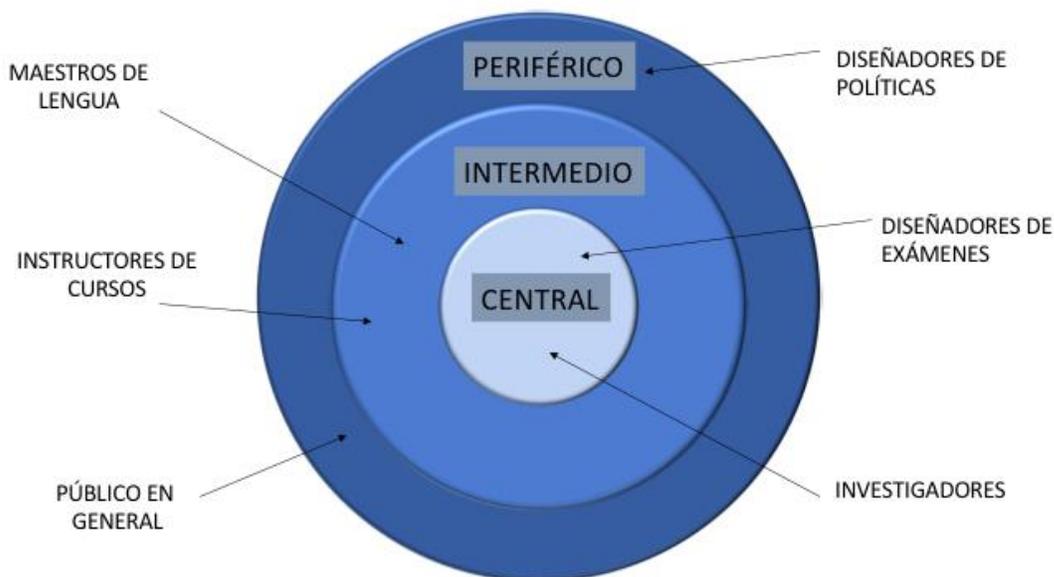
1.2.3.2 Alcance de la literacidad en evaluación de profesores de lengua

El proceso de diseño de exámenes involucra a varias instancias, como se ha comentado en este trabajo, que inicia con aquellos interesados en conocer un nivel de competencia lingüística acerca de un individuo en particular, candidato o estudiante, los administradores de los procesos y de las instituciones, los evaluadores, los diseñadores de exámenes, los profesores y por supuesto, dichos candidatos o estudiantes. Generalmente son los

profesores de lengua quienes diseñan los exámenes, aunque también hay profesionales de evaluación, quienes con frecuencia son también profesores de lengua, y que se dedican a desarrollar reactivos para exámenes estandarizados o a investigar sobre los procesos de evaluación y diseño de exámenes. Taylor (2013) explica quiénes son los involucrados en el proceso de diseño de exámenes: coloca en un plano central o nuclear, a los diseñadores de exámenes y a los investigadores; en un segundo plano, que denomina intermedio, a los profesores y a los instructores de cursos; y en un tercer plano, llamado periférico, a los diseñadores de políticas y al público en general. Los ubicados en el centro son quienes más competentes debieran ser en evaluación de lenguas (Figura 1.1):

Figura 1.1.

Niveles diferenciados de literacidad en evaluación de lenguas de acuerdo con el sector de involucramiento



(Taylor, 2013:409).

En los estudios sobre literacidad en evaluación de lengua se percibe notablemente la necesidad de hacer la distinción entre lo que deben saber o las competencias que deben tener los profesionales de evaluación de lengua y los profesores (Inbar-Lourie, 2008; Fulcher, 2012; Jeeong, 2013; Vogt y Tzagari, 2014; y Giraldo, 2017). Inbar-Lourie (2008)

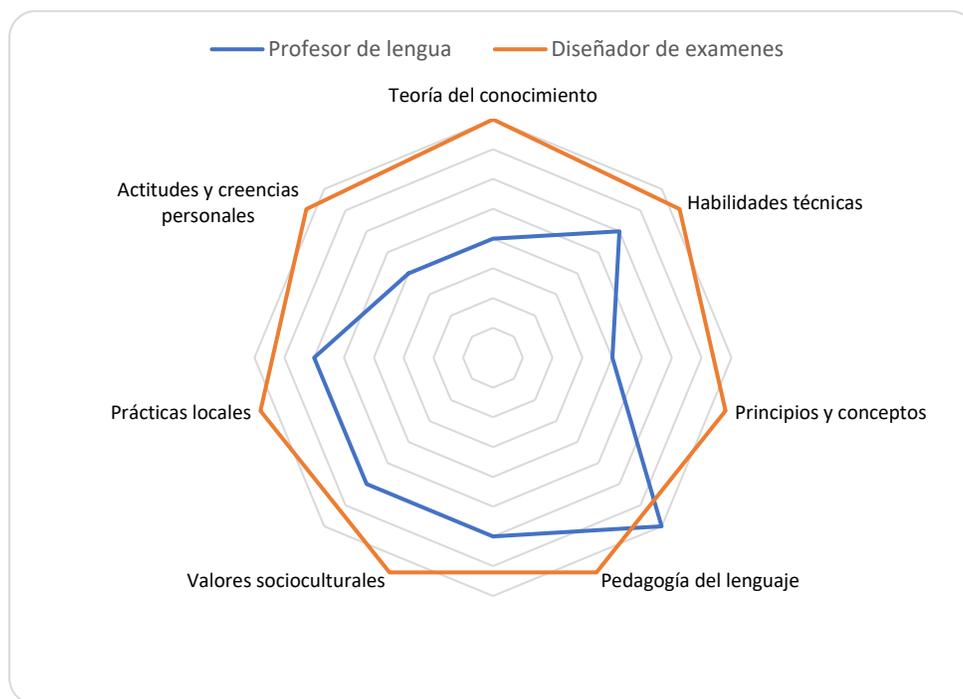
menciona que hay una brecha entre ambos, debido a perspectivas epistemológicas diferentes, una visión conductista y altamente especializada de los diseñadores de evaluaciones contra un enfoque constructivista del profesor, cuyo enfoque es principalmente en el aprendizaje.

Taylor (2013:410) muestra gráficamente cómo el profesional en evaluación debe tener un perfil que cumpla con un conocimiento de la teoría, habilidades técnicas, principios y conceptos, pedagogía en enseñanza lengua, conocer sobre prácticas locales, conocer los valores socioculturales, tener ciertas creencias y actitudes personales, y sobre resultados y toma de decisiones.

El trabajo del profesional de diseño de evaluaciones y exámenes de lengua y del profesor de lengua, entre cuyas funciones se encuentra precisamente, diseñar evaluaciones y exámenes, no son tan diferentes. Sin embargo, el grado de literacidad en evaluación sí es distinto (Taylor, 2013). En la figura 1.2 se comparan los dos perfiles de literacidad requerida en el diseñador de exámenes profesional y el profesor de lengua:

Figura 1.2.

Perfil de literacidad en evaluación del profesor y del diseñador de exámenes de lengua.



Nota: Basado en Taylor (2013, p.410).

El perfil que presenta Taylor del profesor de lengua es diferente al del profesional de evaluación en que el nivel más alto de conocimiento es únicamente en un aspecto, la pedagogía de la lengua, pero es muy poco lo que se le exige al profesor en cuanto a la teoría de diseño de exámenes, principios y conceptos, interpretación de resultados y toma de decisiones.

En el caso de los profesores de la universidad que se estudia en esta investigación y con base en la observación y conocimiento del contexto por parte del investigador, es posible afirmar que el nivel de literacidad coincide al descrito como intermedio por Taylor (2013) en la figura 1.1, que es superior al del público en general, pero inferior al de investigadores especialistas en el área. Por otro lado, es necesario realizar estudios que permitan definir un perfil de literacidad con base en los factores presentados por Taylor (2013) en la figura 1.2.

Davies (2008) y Taylor (2009) puntualizan la necesidad de la profesionalización de los responsables del diseño de exámenes de lengua y de un profundo conocimiento y entendimiento por parte del resto de los involucrados, debido a la importancia que tienen en distinto ámbitos, como se ha mencionado a lo largo de este trabajo.

La evaluación es un aspecto inherente a la docencia. En el caso de la evaluación de lenguas, la literatura muestra que el perfil que se requiere de un profesional de evaluación es más especializado que lo que se exige al profesor de lengua (Taylor, 2013). No obstante, dada la importancia de la evaluación en los sistemas educativos formales, entendiéndose niveles educativos, es necesario que el profesor busque mecanismos de compensación para dicha insuficiencia y de ahí la importancia de la literacidad en evaluación. En el siguiente inciso se presenta la teoría y literatura acerca del diseño de exámenes de lengua, que representa el aspecto técnico en la literacidad de evaluación de lenguas.

1.3. Exámenes de lengua

Los exámenes son el instrumento de evaluación predominante en el ámbito educativo, incluyendo la enseñanza de lenguas. A pesar de la necesidad de adaptación a los paradigmas educativos que rigen actualmente, el diseño de exámenes es un imperativo en la práctica profesional para el lingüista aplicado, el evaluador, o el profesor de lengua. Las razones son diversas; Douglas (2014) argumenta la igualdad de oportunidades o equidad

que deben tener los estudiantes o candidatos y agrega que los exámenes nos proporcionan confiabilidad e igualdad de condiciones, para que los interesados en los resultados de los exámenes (*stakeholders*) puedan fiarse del resultado de un candidato y que se pueda entender a través de algún estándar, como un marco común de referencia. La postura crítica sobre los exámenes de lengua (Shohamy, 2007) argumenta razones políticas que fomentan la continuidad de los tests como instrumento de control por parte de los organismos gubernamentales.

Es así que ya sea como instrumentos de control o estandarización, como medio de garantizar confiabilidad, por tradición, costumbre o resultado de resistencia al cambio, los exámenes son el medio de evaluación que predomina aún en la educación y la enseñanza de segunda lengua (Brown y Abeywickrama, 2018).

He aquí la imperante necesidad para el docente y, en el caso que nos ocupa, el docente de segunda lengua de conocer a profundidad el proceso de diseño y aspectos técnicos, como tipo de tareas o técnicas, características fundamentales de los exámenes de lengua (*test usefulness*), análisis de reactivos y análisis estadísticos. El conocer los aspectos técnicos le da al diseñador bases para la toma de decisiones, sobre todo en busca de lograr validez y confiabilidad en los instrumentos.

En esta siguiente sección del trabajo se revisarán algunos estudios y áreas de estudio en cuanto a diseño de exámenes; se explorarán algunas definiciones y conceptos básicos, y se presenta el modelo de características de un examen (utilidad o aplicabilidad, *test usefulness*) de segunda lengua. Si bien, la literatura es extensa, se busca tratar los temas y conceptos más importantes y así dar al lector una visión sobre los aspectos técnicos del diseño de exámenes de lengua.

1.3.1 Antecedentes de estudios de diseño de exámenes de lengua

Existe una amplia literatura acerca de evaluación y diseño de exámenes de lengua. Los exámenes de lengua han sido motivo de estudio desde diversas perspectivas y enfoques. Se ha estudiado su validez (Messick, 1989, 1996; Kane, 2001; Davies y Elder, 2005; Chapelle, 2011, 2013; Davies, 2012; Velasco Ariza, V., Anguiano Paliza, M. y Larrazolo Reyna, N, 2007; Mendoza Ramos 2015); así como el impacto de los exámenes en los usuarios (Hamp-

Lyons, 2000; Taylor, 2005); el impacto social (Shohamy, 2000, 2001a, 2001b; McNamara, 2001; McNamara y Roever, 2006; el impacto en el salón de clase de segunda lengua (Alderson y Wall, 1993; Cheng, 1997; Green, 2006); por su utilidad o aplicabilidad (Bachman y Palmer, 1996, 2010; Moore, Andrade y Matamoros, 2010); por su grado de equidad, objetividad o justicia (Kunnan, 2000, 2001; Shohamy, 2000, Yin y Wu, 2017); entre muchos otros estudios.

En cuanto a los procesos de diseño de exámenes en particular, han sido descritos o propuestos por diferentes autores a lo largo de los últimos veinte años. Genesee y Upshur (1996), Bachman y Palmer (1996), Hughes (2003) y Fulcher (2013) proponen modelos similares que incluyen como principales etapas: la definición del constructo a medir, la redacción de una tabla de especificaciones, diseño de tareas y reactivos, realización de prototipo o pilotaje, administración del examen y obtención de puntajes o calificaciones. Este proceso es generalmente aceptado, sin embargo, no es posible aplicarlo a todos los contextos, ya que la realidad de los profesores mexicanos es muy distante a lo propuesto en la teoría anglosajona.

Varios esfuerzos se han hecho para resumir el estado del arte en el área de diseño de exámenes de lengua, pero como lo señalan Skehan (1988) y Davies (2013), a partir de fines de los años ochenta se despertó un interés creciente en investigar los distintos componentes de este proceso. Bachman (2000), Spolsky (2008), Farhady (2009) y, con mayor profundidad, Davies (1978, 2013), Skehan (1988, 1989) y Alderson y Banerjee hacen recuentos del desarrollo de la evaluación y del diseño de exámenes de lengua e investigación en estas áreas a través de los años.

De estos escrutinios, podemos resumir y establecer tres etapas del diseño de exámenes, la psicométrica, la comunicativa y la social. La primera etapa, la psicométrica, tiene origen en los sesenta, cuando el diseño de los exámenes de lengua se desarrolla como una actividad de especialización, se tomaron como base la metodología y los instrumentos de medición de constructos psicológicos y se buscaba que estos instrumentos cumplieran con los requerimientos psicométricos. La segunda fue la etapa comunicativa, en los ochenta, cuando el desarrollo del enfoque comunicativo de la enseñanza de lengua y el desarrollo de los modelos de competencia lingüística (Canale y Swain, 1980; Canale 1983) generaron un

mayor enfoque en que los exámenes tomaran aspectos que asemejaran funciones del lenguaje en la vida real. La tercera etapa, la social, en la que la validez, la argumentación del uso del examen y el impacto en las personas involucradas cobran mayor importancia y como predijo Bachman (2000), se volvería importante objeto de estudio de los especialistas en el área.

Después de hacer un recorrido por el desarrollo de exámenes de lengua desde los años sesenta hasta la primera década del siglo XXI, Davies (2013) concluye que es necesario hacer reflexiones en tres sentidos, en cómo examinamos, qué examinamos y quiénes son los examinadores. Davies menciona que ha habido desarrollo en cuanto a cómo construimos los exámenes de lengua, con técnicas más precisas, con mayor atención en diversos aspectos técnicos y métodos de análisis; también en cuanto a conocimiento de la lengua ha habido avance, sin embargo, concluye Davies, este conocimiento de la complejidad del constructo de la competencia lingüística nos ha llevado a cuestionar, paradójicamente, el profesionalismo de los especialistas de lengua; lo que nos lleva a una reflexión más, mientras se ha perfeccionado la técnica y la definición de los constructos o habilidades a medir y se ha fortalecido el desarrollo de especialistas en el área y la profesionalización de los examinadores, también crece la necesidad de fortalecer el aspecto ético, de rendición de cuentas, de literacidad en evaluación y de seguir investigando tres aspectos claves y sensibles de la evaluación y diseño de exámenes que son el retroefecto (*washback*), validez y ética.

1.3.2 Definición y características de los exámenes de lengua

Genesee y Upshur (1996, p.141) definen examen como “*a set of standard tasks that are attempted by a group of students in order to demonstrate their knowledge or ability*” (Un conjunto de tareas a realizar por un grupo de estudiantes para demostrar su conocimiento o habilidad³). En el caso de los exámenes de lengua los candidatos tratan de demostrar su habilidad lingüística. De acuerdo con estos autores los tests o exámenes tienen tres atributos: contenido, método y medida. El contenido se refiere a que el examen es acerca de ‘*algo*’ (Genesee y Upshur, 1996, p.141) en particular; en segundo lugar, el método, cada

³ Traducción propia

tarea representa una forma o procedimiento de obtener información sobre ese *algo*; y en tercer lugar, con base en un marco de referencia estos métodos obtienen información acerca de una característica de un individuo.

Estos atributos se relacionan de la siguiente manera ejemplificada: Con la finalidad de conocer acerca de una característica de un individuo, como su habilidad lingüística en inglés como segunda lengua (contenido), se toma como base un marco de referencia, por ejemplo, el Marco Común de Referencia Europeo, para interpretar un resultado o calificación (medida) obtenidos a través de la aplicación de tareas (método) que al ser desempeñadas por el candidato manifiesta esa habilidad o constructo.

Hughes (2003, p.8) sugiere que los exámenes de lengua o sistemas de examinación deben proporcionar de manera precisa una medición de la característica en particular en la que se tiene interés, que debe tener un efecto benéfico para el proceso de enseñanza de lengua y que esté al alcance en términos de tiempo y monetarios. Además de estas características, Hughes enfatiza la importancia de tener mucha claridad en cuanto al propósito de un examen. Un examen podría estar diseñado bajo una serie de especificaciones muy rigurosas y aun así no ser válido si no se utiliza para el propósito con que fue diseñado. Es común ver que exámenes de certificación estandarizados, como los mencionados de Cambridge Assessment o de ETS, sean utilizados para medir el logro del estudiante en un curso o como diagnóstico previo al mismo, cuando esos instrumentos no fueron desarrollados para tal propósito y por consecuencia su uso en ese sentido es cuestionable.

1.3.3 Utilidad de un examen de lengua (*Test usefulness*).

La teoría presentada por los autores mencionados expresa la práctica deseable en cuanto a diseño, selección y aplicación de exámenes de lengua. Todos los autores enfocan sus esfuerzos a la realización de exámenes válidos y confiables; Bachman y Palmer (1996), además, proponen la evaluación del examen mediante un análisis de su utilidad (*test usefulness*), que se compone de seis cualidades o principios: validez del constructo, confiabilidad, practicidad, autenticidad, interactividad e impacto. En la traducción al español del término *usefulness*, se ha utilizado la palabra ‘utilidad’ (Moore, Andrade y Matamoros, 2010), sin embargo, consideramos que el término ‘aplicabilidad’ sería más adecuado a lo que Bachman y Palmer pretenden, determinar a qué grado es aplicable un test

en un contexto determinado, y si el término ‘utilidad’ bien podría cumplir con dicha función, consideramos que una connotación negativa podría implicar desechar o invalidar un examen, cuando solo se debería considerar a qué grado es aplicable.

Enseguida se analizan los componentes del concepto de utilidad o aplicabilidad de examen de lengua (*test usefulness*).

1.3.3.1 Validez

En cuanto a componentes de diseño de exámenes se refiere, el término de validez es el más importante y es el objeto de mayor número de estudios por parte de investigadores en el área (Davies, 2013). Un instrumento de medición será válido mientras mida lo que se pretende medir. Esta definición, generalmente aceptada, tiene muchos matices y genera muchos cuestionamientos ya que el término en sí es difícil de definir por todas sus aristas e implicaciones. El desarrollo de esta característica se resume brevemente en los siguientes párrafos.

El término de validez, y en particular la validez del constructo, es un término propuesto para estudios psicológicos por Cronbach y Meehl (1955), quienes plantean que la operacionalización del constructo a medir es un requisito indispensable para validar un test; en otras palabras, para cada prueba habrá que desarrollar una definición teórica para el constructo que se pretende estudiar. Sin embargo, Messick (1989), plantea que no es suficiente con definir el constructo, sino que para validarlo es preciso considerar las consecuencias sociales que el resultado del examen tendrá en la vida del candidato, considerando las decisiones que se tomarán con base en ese resultado.

Dentro del concepto de utilidad, el término validez del constructo cobija todos los tipos de validez. Bachman y Palmer (1996, p.21) definen validez del constructo como “el grado al que podemos interpretar un resultado dado como indicador de las habilidades o el constructo que queremos medir⁴”. Esto implica, de acuerdo a Bachman y Palmer, que no es suficiente con argumentar que un instrumento es válido, sino que es necesario construir y

⁴ Traducción propia.

mostrar evidencia de que el instrumento mide el constructo y si debe ser medido de esa manera.

Este planteamiento de Bachman y Palmer deriva de lo establecido por Messick (1989, p.13), quien sugiere:

Validez es un juicio de evaluación integrado acerca del grado en que una evidencia empírica o argumento teórico respalda la suficiencia, pertinencia e idoneidad de las inferencias y acciones basadas en los resultados de los exámenes y otros tipos de evaluación.⁵

Así que, de acuerdo con esta definición, el objeto de la validez debe ser justificar y clarificar las interpretaciones que se hagan y el uso que se le dé a un resultado de un examen o evaluación (Kane, 2001). Una corriente de estudio y de pensamiento surge a partir del planteamiento de Messick (Kane, 1992; Bachman y Palmer, 1996 y 2010; McNamara, 2000; Weir, 2005; McNamara y Roever, 2006; Chapelle, 2011; Chalhoub-Deville, 2016), buscando que los exámenes de lengua cuenten con esa argumentación o justificación y la consideración del impacto del test durante su desarrollo. Dada la complejidad del constructo que nos ocupa en la evaluación de lengua, la competencia comunicativa, el planteamiento de los autores debería ser de la mayor importancia, lo cual en la práctica no sucede, como se verá más adelante en este estudio.

Davies (2013) pone sobre la mesa la pregunta si la validez de un instrumento está en el mismo instrumento o en sus consecuencias. Hay diferentes posturas en este sentido y es que un examen puede ser válido desde el punto de vista psicométrico, cumplir con medir el constructo para el que fue diseñado o si el uso que se le da al examen corresponde con su diseño original. Para Messick (1989) y Bachman y Palmer (2010) la validez de constructo es un término que agrupa el resto de los tipos de validez, así que el diseñador de exámenes debe procurar y buscar que el instrumento sea válido, en estas dimensiones y tomar responsabilidad por los instrumentos que diseña y considerar tanto medidas psicométricas y aspectos técnicos del diseño, como buscar la manera que realmente mida la competencia comunicativa en alguna de sus manifestaciones o componentes.

⁵ Traducción propia.

Schwandt y Jang (2004) debaten entre el enfoque técnico de medición de lenguaje por medio de instrumentos psicométricos, donde la estadística juega un papel crucial para definir si un instrumento es válido o no, contra el pensamiento o enfoque acerca de la ética y el uso que se dé a los exámenes de lengua, considerando la complejidad del constructo. Los autores proponen los términos *Scientific Practice (Práctica científica)* y *Everyday Practice (Práctica diaria)* para diferenciar dos enfoques de diseño de exámenes, el técnico y el político. El primero parece prevalecer sobre el segundo debido a la dificultad que representa medir la lengua en su contexto, en su caótico uso real, sin embargo, la práctica diaria, es donde el profesor interactúa con el candidato o estudiante, el contexto lo envuelve y lo lleva a tomar decisiones que generalmente parecen determinar el grado de validez del instrumento.

La validez es la cualidad más importante con que debe contar un instrumento de medición de lengua, y si un examen no es válido, su uso o las decisiones que se toman con base en él no son éticas y no se es responsable si no se actúa bajo una ética. Este estudio busca entender, bajo estos supuestos, las razones o justificaciones que tienen los profesores para su actuar, identificar como expresan su responsabilidad profesional, cualidad que conlleva un importante componente ético, que a su vez se relaciona estrechamente con la validez.

1.3.3.2 Confiabilidad

Otro concepto esencial en el diseño de exámenes es la confiabilidad o fiabilidad. Los exámenes de lengua deben mostrar una consistencia en los resultados si un candidato o estudiante tomara un examen en dos ocasiones diferentes bajo circunstancias similares (Fulcher, 2007, p.104), o como lo especifican Bachman y Palmer (1996) una consistencia en la medición. Esta cualidad de los *tests* implica factores del propio examen y factores externos, ya que si en su diseño la prueba incluye tareas que no replican los resultados obtenidos por tareas similares en otro test de condiciones parecidas el instrumento no es confiable. Por otro lado, si en su administración el examen implica irregularidades de factor externo, es decir ajenas a su diseño, como ruido, interrupciones o cuestiones físicas del examen, el instrumento también verá afectada su confiabilidad puesto que los resultados de dos exámenes serían incomparables y las inferencias sobre la habilidad del candidato o alumno serían erróneas.

Aunque en un contexto diferente al caso que se estudia en este trabajo, un ejemplo de la confiabilidad de exámenes de lengua es los exámenes realizados por Cambridge Assessment o por ETS, que solicitan a sus centros examinadores una serie de estrictas medidas en cuanto a la aplicación de los exámenes con el fin de asegurar la confiabilidad de resultados. Es común en el mundo laboral y académico que a las personas se les solicita obtener alguna certificación emitida por estas instituciones que se basan en el resultado del examen. El interesado en el resultado (*stakeholder*) debe tomar una decisión con base en ello y por eso solicita estas certificaciones que otorgan resultados confiables, aun así, no en todos los casos son válidos.

Lo anterior podría ejemplificarse con el caso de una compañía que solicita a sus candidatos a ocupar posiciones dentro de la empresa evidencia de competencia en lengua inglesa y para ello les pide el certificado emitido después de tomar el *Test of English as a Foreign Language* (TOEFL) desarrollado por ETS. Este examen está diseñado para los candidatos a ocupar un lugar en una universidad donde los estudios se llevan a cabo en lengua inglesa. Los requerimientos de ETS a sus centros aplicadores aseguran que el examen es confiable, no obstante, su validez para los candidatos a ocupar un puesto en la empresa es cuestionable, ya que la competencia es diferente, el uso de la lengua que se puso a prueba es distinto al fin que se le iba a dar.

En un segundo ejemplo, el caso de candidatos a estudiar un posgrado en una universidad, a quienes se les solicita una certificación de comprensión de textos emitida por la misma universidad a aquellos candidatos que aprueben el examen. Si por alguna razón para obtener evidencia de la comprensión lectora en lengua inglesa de los candidatos un instrumento es utilizado repetidamente y los candidatos notan esto, las respuestas del test podrían darse a conocer lo que afectaría la confiabilidad del examen. Los interesados en el resultado no podrían confiar en esta información, por lo tanto, el examen no solo no es confiable, sino que tampoco puede ser válido para los propósitos para los que se usó o diseñó. De estos dos ejemplos se puede concluir, que si un examen no es confiable, tampoco es válido; pero que un examen sea confiable no implica que sea válido, como lo expresa Davies (2011, p.38) “la confiabilidad le da forma a un test, la validez le da significado”; consecuentemente, la confiabilidad es una cualidad esencial para obtener validez en un instrumento.

Hughes (2004) trata de manera amplia los aspectos a considerar en el diseño y aplicación de un examen de lengua que pudiese tener un impacto en su confiabilidad. Hughes tiene un enfoque muy claro en los exámenes estandarizados, de certificación o a gran escala, como son los de Cambridge Assessment, ETS o Michigan Assessment, no obstante, varios de estos aspectos son aplicables en el caso de los exámenes en el salón de clase de lengua.

De acuerdo con Hughes (2004) la muestra de competencia lingüística o comunicativa debe ser representativa, los exámenes no pueden ser muy cortos, ni su contenido. Si la muestra pretende ser significativa, deberá ser amplia. En caso de no ser así, se corre el riesgo de que el resultado no sea representativo de la conducta que se busca producir en el alumno.

Los exámenes deberán ser conformados en su gran mayoría por reactivos que ayuden a hacer una distinción entre los alumnos más fuertes y los más débiles en cuanto a su competencia comunicativa. Los reactivos muy fáciles y los que son muy difíciles no contribuyen a obtener un resultado confiable.

Hughes (2003) recomienda no otorgar mucha libertad a los candidatos en las tareas. Esto es, si los alumnos reciben instrucciones muy generales sobre cómo realizar una tarea, el desempeño sería muy variable y de igual manera los resultados. Si los alumnos reciben instrucciones precisas y específicas que no les permitan una variedad de respuestas el resultado será más confiable.

Para lograr la confiabilidad de un examen es esencial que los reactivos no sean ambiguos, sino que deben ser muy claros en la pregunta o raíz y en las respuestas u opciones de respuesta, para que el alumno pueda responder el reactivo con asertividad y el profesor esperar una respuesta en concreto y no hacer consideraciones para cada respuesta encontrada. En este caso el pilotaje, a través de la revisión de colegas, o de una aplicación previa, permite identificar reactivos con respuestas ambiguas. Hughes también recomienda contar con una clave de respuestas que incluya otras posibles respuestas de parte del alumno y una decisión a priori sobre cómo calificar las respuestas alternativas.

Las instrucciones de las tareas a realizar en el examen deben ser claras, de manera que los estudiantes sepan con precisión qué deben de hacer y no haya lugar a interpretaciones. Para

evitar esto, es recomendable realizar prácticas de las tareas a realizar en un examen, de manera que el alumno esté familiarizado anticipadamente.

Los exámenes deben ser perfectamente legibles, con espacio suficiente para contener las respuestas del alumno y una estructura y organización que permita que el alumno se enfoque exclusivamente en la resolución del examen y en ningún otro aspecto relacionado con la apariencia del examen.

Durante la aplicación de un examen es recomendable evitar y prever factores que pudiesen causar distracciones a los alumnos, como ruido, movimiento o interrupciones. Hughes recomienda poner énfasis en la uniformidad de la aplicación, es decir que todos los alumnos respondan el examen bajo las mismas condiciones.

Hughes recomienda que la calificación sea objetiva para lo cual el formato de opción múltiple y sus variedades contribuye significativamente. En el caso de los exámenes de habilidades productivas, contar con una rúbrica que establezca claramente los criterios a evaluar y la calificación precisa por el cumplimiento de un criterio, y de esta manera no dejar lugar a subjetividad o criterios variables de parte de los profesores.

Estas medidas contribuyen a incrementar el nivel de confiabilidad de un examen. Como puede apreciarse, estas consideraciones deben ser tomadas en cuenta desde el proceso de diseño del examen, pilotaje, reproducción, aplicación y calificación, y considera a los participantes en el proceso, profesores-diseñadores del examen, aplicadores, profesores-evaluadores y estudiantes.

1.3.3.3 Practicidad

La tercera cualidad de la utilidad de exámenes es la practicidad que implica que los recursos invertidos en el diseño, aplicación y calificación de un examen no excedan los recursos disponibles (Bachman y Palmer, 1996, p.35). De acuerdo con Hughes (2003, p.56) un examen debe ser sencillo y su diseño, aplicación, calificación e interpretación no deben ser costosos. Un ejemplo de practicidad son los exámenes en los que se opta por utilizar reactivos de opción múltiple que son rápidos y fáciles de calificar, aunque en su diseño no lo son, si es así en el resto de las fases mencionadas.

Los beneficios de la practicidad de los exámenes de lengua son muchos y muchos los actores que intervienen en el proceso que son favorecidos. No obstante, es necesario que siempre se privilegien la validez y confiabilidad de los exámenes. Resulta muy sencillo que las personas que intervienen en el proceso de examinación se dejen llevar por las ventajas de realizar exámenes prácticos, que en ocasiones podrían no ser tan válidos o no tan confiables. El aspecto de validez debe de ser procurado en todo momento, así que la tarea del diseñador de exámenes es buscar la manera de replicar situaciones de uso de la lengua o tareas que permitan un uso e interpretación de los resultados válidos de manera que el candidato encuentre justa su evaluación.

En la práctica esta situación no es tan sencilla. Los diseñadores de exámenes deben de cumplir con los requerimientos de las instituciones o las personas a cargo de la administración del examen y las limitaciones en cuanto a tiempo y recursos humanos, financieros y materiales son claramente establecidas. Esto pone al diseñador de exámenes en una posición complicada puesto que debe cumplir con esos límites y con el candidato, desarrollando un instrumento justo. Lamentablemente en muchas ocasiones en la práctica se privilegia la practicidad sobre la validez y la confiabilidad, lo que va en contra de los principios fundamentales de una evaluación.

1.3.3.4 Autenticidad

Las tres primeras cualidades que conforman la utilidad del examen son características deseables en exámenes de cualquier área del conocimiento. La autenticidad, cuarta cualidad, es una característica que es especialmente importante en el diseño de exámenes de lengua. La autenticidad es el grado en que el uso de la lengua en la tarea a realizar en el examen se asemeja un uso real de la lengua (Bachman y Palmer, 1996, p.23). Por ejemplo, pedirle al candidato o al alumno llevar a cabo un *role play* en el que una persona solicita a otra informes o instrucciones para llegar a un lugar, es un típico ejemplo de un uso real de la L2. En diversos contextos puede pedírsele a los candidatos que realicen diversas actividades o apliquen diversas funciones de la lengua como presentaciones, entrevistas, escritos como cartas, correos, ensayos, obtener información de un audio o de una lectura, todas las anteriores actividades que llevamos a cabo día con día en la vida real y es tarea del examinador traer estas situaciones al aula o simularlas en el proceso de evaluación.

La importancia de la autenticidad radica en que su esencia la relaciona estrechamente con la validez del constructo y con la obligación del examinador a relacionar el constructo con el uso real de la lengua. Es así, que considerar la autenticidad de un examen de lengua como otro tipo o subcategoría de la validez del constructo no estaría fuera de lugar.

1.3.3.5 Interactividad

La siguiente característica es la interactividad, que de acuerdo con Bachman y Palmer (1996) es el grado en el que las características del candidato contribuyen a que se involucre en la tarea a realizar en el examen. Si el candidato o alumno a ser evaluado puede lograr una conexión con su conocimiento del mundo, activa sus esquemas, lo que le permite un mejor entendimiento de los textos y por ende un mejor resultado. Como ejemplo, supongamos que en un examen se incluye un texto para evaluar la comprensión de lectura, el tema del texto es un análisis sobre las políticas económicas implantadas por el gobierno. Este mismo texto puede tener un efecto muy diferente en personas con perfiles distintos, pongamos el caso de un economista y de un estudiante de preparatoria, el primero podría fácilmente entender el texto ya que sus esquemas se activarían y relacionaría el texto con su experiencia o conocimientos previos del mundo. Para el segundo candidato, el estudiante, el esfuerzo por entender el texto sería mucho mayor dado que su conocimiento del mundo y del tema no es como el del economista. En esta situación y bajo un supuesto que el nivel de competencia comunicativa sea similar en ambos individuos se esperaría un mejor resultado de parte del economista debido a su conocimiento del tema, lo cual indica que éste es un factor relevante en el resultado del test.

Es responsabilidad del diseñador de exámenes que en medida de lo posible esto no suceda o que el impacto del tema en el resultado que obtenga el individuo sea el mínimo. Es inevitable que los temas de los textos influyan en los resultados de los candidatos, lo que si habrá que evitar es que se vean perjudicados y buscar que ese impacto sea positivo en lugar de negativo.

1.3.3.6 Impacto

La última cualidad es el impacto que puede tener el examen o su resultado en la vida del alumno o el candidato. Este tema se toca de manera repetida a lo largo de este trabajo, ya

que es esencial en el aspecto ético del profesional de diseño de exámenes. La literatura acerca del impacto de los exámenes de lengua, específicamente desde el punto de vista crítico (Shohamy, 2000, 2007; McNamara y Roever, 2006; Bachman y Palmer, 2010). La literatura sobre evaluación de lengua en el salón de clase ha tenido dos vertientes, un enfoque en el desarrollo de argumentación de la validez en los exámenes (Bachman y Damböck, 2108) y las más común, que es el impacto de los exámenes en la instrucción, en la práctica docente, lo que se le conoce como *washback* o retroefecto (Hughes, 2003; Brown y Abeywickrama, 2018).

De acuerdo con Hughes (2003), el *washback* puede ser positivo o negativo. Será el primer caso cuando los exámenes promuevan el aprendizaje y el logro de los objetivos pedagógicos y que durante un curso contribuyan al desarrollo de competencia que después sería puesta a prueba en el examen. El efecto *washback* será negativo, cuando la instrucción desvíe el curso del logro de los objetivos, cuando la preparación para el examen se ubique por delante del aprendizaje.

Al impacto también se le conoce como validez consecuencial (Brown y Abeywickrama, 2018), e implica la consideración del alumno ante el examen y prácticas como dar a conocer el formato, preparar a los alumnos, eliminar sesgos y promover el aprendizaje mediante el instrumento.

Conclusión al capítulo

En este capítulo se ha presentado de manera deductiva el tema central de la investigación: La evaluación del aprendizaje de lenguas. Este se enmarca en un ámbito general, la evaluación educativa para transitar hacia un marco particular, la evaluación del aprendizaje de lenguas, y dentro de este tema, se particulariza en los exámenes, que son el instrumento de más uso entre los docentes en el estudio de caso. Se ha buscado explicar la visión que se tiene con respecto a la evaluación del aprendizaje, desde el paradigma epistemológico de la complejidad, así como las diferentes posturas y enfoques existentes entre la literatura especializada.

Se ha hecho una revisión de la literatura que se enfoca en el trabajo del profesor en el salón de clase y se han explorado los aspectos técnicos relativos al diseño de exámenes que

prevalecen en la teoría, que después serán confrontados con la práctica real de los participantes.

Todos los aspectos mencionados, relacionados con la evaluación, convergen en la competencia denominada literacidad en evaluación de lenguas, que contribuye a la construcción del perfil profesional del evaluador y como referente de la práctica del docente en ese ámbito de su práctica profesional. Asimismo, los componentes de la literacidad en evaluación constituyen una base para el entendimiento de la práctica evaluativa del profesor de lengua y de los hallazgos de este trabajo que se presentarán más adelante. Los aspectos técnicos de la literacidad se han presentado en este capítulo, utilizando como fundamento el concepto de utilidad (*test usefulness*) y otro de sus componentes, el aspecto ético, se trata en el capítulo siguiente.

CAPÍTULO II

ÉTICA EN LA EDUCACIÓN Y EVALUACIÓN DE LENGUAS

Introducción al capítulo

El profesor de educación superior tiene una función social que conlleva la guía y transmisión de conocimiento y valores (González Maura, 2004), debe ser un modelo para el estudiante y esto no es diferente para el caso en cuestión que son los profesores de lengua. Por lo tanto, se espera que la conducta del profesor y sus acciones se encaminen, además del desarrollo académico y personal del alumno, a una ética de trabajo profesional y personal.

En este capítulo se presenta la conceptualización que estructura el entendimiento de esta ética profesional, enfocada principalmente en el ámbito educativo y particularmente en el docente de lengua y la evaluación del aprendizaje.

La mirada a la ética desde la perspectiva de la complejidad (Morin, 2006) sustenta epistemológicamente los argumentos presentados, así como sirve de punto de partida para un recorrido a través de la ética en la educación, el docente y la evaluación, para concluir con el enfoque particular del caso y del capítulo, la ética en el diseño de exámenes de lengua, donde se exponen las principales aristas y enfoques que rodean este aspecto de la docencia de lengua, como son los enfoques en la validez y en la equidad.

2.1 Ética

La ética es un imperativo, sentencia Morin (2006), una exigencia originada por tres factores, tanto en el interior del individuo, como por factores externos y por una “fuente anterior”, que Morin relaciona con la existencia previa de la especie y que se ha transmitido genéticamente. La ética, por lo tanto, es inherente al ser humano, a lo que él denomina como la triada individuo-sociedad-especie eje del cual emerge la humanidad, define lo humano.

Morin explica que el individuo es un sujeto, quien es el centro de su propio universo o egocentrismo. Este involucra dos principios, el de exclusión y de inclusión. En palabras de

Morin (2006: 22): “El principio de exclusión significa que nadie más que uno mismo puede ocupar el puesto egocéntrico en el que expresamos nuestro Yo”. Por otro lado, el principio de inclusión, habilita al sujeto a que un Nosotros ocupe ese lugar del Yo, dando lugar al altruismo, que desplaza el egocentrismo; como resume Morin (2006, p.23): Ser sujeto es conjugar el egoísmo con el altruismo”, lo cual produce una complejidad.

De acuerdo con Morin, el egoísmo, implícito en el humano rompe lo que él denomina la trinidad individuo-sociedad-especie, que deja de ligarse y la ética y el altruismo son la religación, de éstos, lo que da lugar a las normas que la sociedad busca generar en sus individuos un comportamiento solidario y crear comunidad.

Todo lo anterior sobre la complejidad de la ética de Morin puede entenderse en los dos postulados siguientes:

- “Toda mirada sobre la ética debe reconocer el carácter vital del egocentrismo así como la potencialidad fundamental del desarrollo del altruismo” (Morin, 2006, p.23).
- “Toda mirada sobre la ética debe percibir que el acto moral es un acto individual de religación: religación con el prójimo, religación con una comunidad, religación con una sociedad y, en el límite, religación con la especie humana” (Morin, 2006, p.24)

Por consiguiente, se entiende por ético aquello que coadyuva en la religación de la trinidad y que contribuye a la comunidad, a través del comportamiento individual.

En el caso de estudio se consideran estos postulados éticos y se busca entender el proceso de evaluación, los individuos participantes y sus acciones y decisiones. Cabe la aclaración que no se hacen juicios éticos, sino que la ética en la complejidad de Morin solo es la base del entendimiento de la ética, que es un factor importante en el desempeño profesional.

2.1.1 Concepto de ética

La labor de los docentes de lenguas se rige por los principios de cada profesional, de cada individuo y a estos principios se les conoce como moral. De acuerdo a Martínez Navarro (2010:24) la moral es un “conjunto de creencias, preceptos y prácticas acerca de lo que se considera como comportamiento humanamente correcto en un contexto espacio-temporal

determinado”. Domínguez Prieto (2003) argumenta que la moral se constituye de un sistema que, a su vez, se integra de “un conjunto de normas, patrones de conducta, mandatos, valores y prescripciones” (Domínguez Prieto, 2003, p.14-15) que el individuo recibe de parte de otros y con ello se desenvuelve en la vida.

El estudio o la reflexión sobre la moral es la ética (Martínez Navarro, 2010). Román (2003:31) menciona que la ética “alude explícitamente a la filosofía moral, a la reflexión crítico-racional sobre las morales, con el objetivo de que éstas sean auténticamente morales”. Domínguez Prieto (2003) coincide con lo anterior y agrega que la ética explica los comportamientos a través de ciertos criterios y los aplica a situaciones prácticas. Ibarra (2005) considera que la ética es “expresión de una conciencia moral que posibilita el logro de un bienestar social y contribuye a la realización plena del profesionista”. Ibarra en este sentido incorpora al concepto de ética dos elementos, que son el bienestar social y la realización del profesionista (Ibarra, 2005:4).

Davies (2004) se refiere a lo que es ético como la ‘regla de oro’ y explica el carácter social de la ética de la siguiente manera: “Una creencia, acción o decisión ética es aquella que esté basada en una convicción de que lo que se hace es lo correcto en el momento en que se hace. La justificación debe ser específica; el interés propio no es suficiente.” (Davies, 2004, p.98). Al tener que decidir qué es lo correcto para actuar no debe pensarse en el bien individual únicamente, sino más bien en el bienestar común de la sociedad. Los individuos por sí solos no podrían lograr muchas cosas y para conseguirlas en muchas ocasiones es necesario que la conducta del individuo se restrinja, por ende, la ética debiera ser universal.

De esta manera, la ética y el profesionalismo están estrechamente ligados debido a su función social. Davies (2004) profundiza: para los profesionales es una necesidad creciente el estar regulados por una moral, ya que eso es lo que los categoriza como tales, la existencia de una serie de principios, generalmente registrados en un código, así como lo menciona Johnstone (2003:9): “En áreas como la docencia y la investigación, el término ética ha tenido un significado más específico, se refiere a estándares codificados y reglas que gobiernan la práctica profesional”. En este sentido Ibarra (2005) menciona que la ética profesional va más allá de un código, de prohibiciones o deberes, o especulaciones filosóficas y hace referencia al carácter práctico de la ética, ya que ésta debe ser una guía

que nos lleve a tomar decisiones y que genere acciones positivas en beneficio de la sociedad y del profesionista.

En resumen, la ética se percibe como un sistema individual, pero relacionado con el entorno, con base en el que el individuo tomaría decisiones de carácter personal, y profesional en este caso, que podrían acercarse a lo “correcto”. Este sistema se conformaría, en concordancia con Domínguez Prieto (2003), por un conjunto de normas, patrones de conducta, mandatos, valores y prescripciones. Esta visión de la ética, nos permitirá más adelante un acercamiento a la comprensión de la conducta del profesor que es objeto de estudio de este trabajo, así como el fundamento de sus decisiones en la práctica profesional.

2.2 Ética en la educación

Una vez señalada la importancia de la ética en el quehacer del profesional en general, se profundizará en el área docente. Hablar de la ética del docente implica relacionar su labor con los principios que la rigen. Es así que se tendría que hablar de la labor del docente, en especial a nivel superior, y definir cómo es un ‘buen’ docente. Ferreyra y Rúa (2018) explican que las prácticas de enseñanza tiene tres dimensiones: la didáctica, teórica y ética. La dimensión ética se refiere a la reflexión del docente sobre qué es una buena enseñanza y “el compromiso por realizar la tarea del mejor modo posible, acortando la distancia entre lo que se hace y lo que se consideraría correcto hacer” (Ferreyra y Rúa, 2018, p.146). En los siguientes párrafos, se busca guiar al lector a conseguir una visión de lo que está implicado en esta dimensión ética de la enseñanza específicamente en el ámbito educativo que nos ocupa, el universitario.

2.2.1 Alteridad

Dentro de los conceptos que rodean la ética, se encuentra el de alteridad, que de acuerdo con Vega (2015) es “cambiar la propia perspectiva por la del ‘otro’, el reconocimiento de las singularidades del otro”. González Silva (2009) menciona el carácter político y ético de la educación, lo que implica un compromiso a “hacerse cargo del otro”, de manera que el alumno reconozca su singularidad es responsabilidad del profesor. Para Ortega Ruíz (2004) la educación no es una relación “profesoral-técnica” entre el profesor y el alumno, sino que es una relación ética, en la que “no es posible educar sin el reconocimiento del otro”

(Ortega Ruíz, 2004, p.10). Para Emmanuel Lévinas (En Quiróz y Marín, 2013) reconocer al Otro es aceptarlo, acogerlo y enfrentarlo y este acercamiento le da un sentido ético. El sujeto requiere de la existencia del Otro, porque, como explican Quiróz y Marín (2013), las soberanías individuales no construyen una moral colectiva, esta se da en comunidad, en el acercamiento al Otro, en su reconocimiento, es decir, en la alteridad.

Con la finalidad de estudiar la alteridad en la educación, González Silva (2009) reportó entrevistas a 45 estudiantes de educación primaria y analizó su discurso, lo que resultó en varias categorías para su estudio: La alteridad puede manifestarse de dos formas de acuerdo con el autor, Alterada o desvinculante y equilibrada o próxima. La primera se clasifica en discriminación por descalificación; tensión discriminación por rechazo o exclusión, empatía (De acuerdo con González de Rivera (2004), “proceso mental voluntario de exclusión de sentimientos, actitudes, pensamientos y motivaciones inducidas por otro”), cosificación, utilitarismo y autarquía. La segunda se clasifica en: tolerancia, empatía, reciprocidad, responsabilidad, y encuentro.

En este trabajo de investigación se comparte la noción de la alteridad como un componente necesario en la educación y por lo tanto en la práctica evaluativa. El carácter social de la educación y los valores como la igualdad y la reciprocidad (House, 1997) se consideran como un elemento que conlleva responsabilidad, que exige al profesional, en este caso, al profesional de la docencia.

2.2.2 El docente universitario

Al referirse a la ética del docente universitario se habrán de contemplar cuatro vectores, de acuerdo a Román (2003): El alumno, el docente, la institución universitaria y la sociedad; y dado el alcance de este proyecto, profundizaremos en los primeros dos: El alumno es el principal vector, y en ese sentido hay que considerar tres situaciones: Román (2003) plantea una relación asimétrica en la que el alumno depende del profesor, lo que implica una responsabilidad de parte de éste; en segundo término, el alumno debe ser consultado, para la toma de decisiones, considerando la naturaleza inequitativa de esta relación; y como tercer factor, contemplar el hecho de que la ética en la educación profesional no deberá basarse en una asignatura en la que se enseñe la ética y moral, sino que debe de tomar un carácter transversal y ser contemplada en cada una de las asignaturas. La autora concluye

enfaticando el hecho que el alumno espera del docente una “formación moral inherente al ejercicio responsable de la actividad para la que sus estudios universitarios deben prepararle” (Román, 2003:50). Como docente es necesario reflexionar en lo expuesto por la autora con respecto al alumno, ya que como se menciona, existe una vulnerabilidad del alumno en la relación con el docente y la institución y esto conlleva una responsabilidad para estos dos últimos.

En cuanto al docente, Román (2003) puntualiza que no es posible separar la moral personal de la moral docente, por ende, un desarrollo integral es fundamental en el ejercicio de la práctica docente. Para Román (2003), el ser un profesional de la docencia implica desempeñarse “correctamente, sin negligencia” (Román, 2003, p.54) y tener presente que el principal objetivo es la formación integral del alumno, en otras palabras, su maduración técnica y humana. La autora agrega (Mauri, 2003, p.162) que al profesional de la educación superior puede exigírsele responsabilidad sobre los contenidos, que deben ser verdaderos, sobre la transmisión de valores éticos y sobre el grado de justicia con que se evalúa a los alumnos.

Ibarra (2018) explica que los profesionales tienen obligación y, como consecuencia, una responsabilidad ante la sociedad, de prestar los servicios que esta requiere. Ibarra se enfoca principalmente en el docente universitario. La autora plantea dos dimensiones de la ética profesional: la ética profesional y la ética específica de la profesión. Define la ética profesional como “un saber filosófico y práctico que ha desarrollado un conjunto de principios y valores que se presentan como guía para orientar la práctica de las profesiones” (Ibarra, 2018, p.74). Por lo tanto, se espera del docente universitario actuar con base en la ética profesional y en la específica de su profesión. En este mismo sentido, Moreno Olivos (2016, p.275) comenta que el ser profesional implica competencias teóricas y prácticas, “integridad personal y una conducta profesional ética”.

En este inciso se han expuesto algunas reflexiones acerca de la ética del docente universitario en sentido social, de su relación con el alumno, del alcance de su responsabilidad, y de su profesionalismo. Más adelante en este trabajo, habrán de analizarse las prácticas de los participantes en el estudio considerando este marco de referencia.

2.2.3 Las actitudes del docente

Martínez Navarro (2010) hace aportaciones en este mismo sentido. Este autor habla de actitudes:

Las actitudes son aquellas predisposiciones, aprendidas y relativamente fijas, que orientan la conducta que previsiblemente se manifestará ante una situación u objeto determinado; la actitud es, pues, una predisposición conducta, que no consideramos innata sino como algo que la persona adquiere y admite modificaciones.” Martínez Navarro (2010:131).

Martínez Navarro (2010:132) relaciona las actitudes con las virtudes, que describe como hábitos de comportamiento deseables, y que contienen componentes cognitivos, afectivos y comportamentales; los cognitivos son sus creencias, convicciones y saberes, los afectivos son los sentimientos, y los comportamentales, las formas en que se responde a ciertos hechos. El autor menciona que para modificar las actitudes se requieren contenidos orientadores que son valores, principios normas e ideales de vida. Las actitudes se internalizan en la persona y conforman su *ethos* moral y profesional, menciona Martínez Navarro (2010), y por ende su actuar se rige por estos y aporta a lo expuesto por Ibarra en que la ética contribuye a la realización del profesionista. Martínez lo expresa de la siguiente manera:

Si el profesor adopta ciertas actitudes éticas, especialmente de actitud básica o fundamental de profesionalidad, de valoración positiva del trabajo bien hecho, entonces vivirá su profesión como un compromiso vital que satisface las aspiraciones más profundas de su persona... (Martínez Navarro, 2010:133).

El autor habla de la reflexión ética como el camino para conocer cuáles son las mejores prácticas, que lleven a una propuesta de ética aplicada a la enseñanza. Esta ética se da en un contexto en el que el docente no actúa en aislamiento sino que tiene una gran injerencia de parte de otros participantes del proceso enseñanza-aprendizaje. Es así como el autor (Martínez Navarro, 2010) describe las relaciones que tiene el docente con estos participantes. Entre las actitudes que debe tener el docente en sus diferentes relaciones el autor menciona las actitudes hacia el alumno. En primer lugar menciona que el docente debe desempeñarse con una actitud de servicio y cooperación con el alumno en lo que requiera para aprender. La segunda actitud mencionada por Martínez Navarro (2010) es de igualdad moral y de respeto. A pesar de la evidente desigualdad en la relación docente-

alumno, ya que este último se encontrará, en la mayoría de las ocasiones en desventaja debido a varios factores, para el docente existe una obligación de no tratar al alumno con desprecio, altanería, manipulación o discriminación arbitraria (Martínez Navarro, 2010:149).

El carácter social de la ética y cómo se da ésta en las relaciones entre el docente y sus alumnos se ha discutido ampliamente (Martínez Navarro, 2010; Román, 2003, Ortega Ruíz, 2004; Petricone, 2006; Cabra-Torres, 2010, Ibarra 2013). Esta relación es un elemento más de todos los que intervienen en el proceso educativo y en particular de evaluación y diseño de exámenes de lengua. Una serie de intereses particulares, que en varias ocasiones se contraponen y hacen de las disyuntivas éticas en la evaluación un problema complejo que debe de entenderse en su entera dimensión.

2.2.4 El docente y la evaluación educativa

Dado el carácter social y, por lo tanto, ético de la práctica docente, es preciso determinar algunos de los alcances de ésta y específicamente en relación a la práctica evaluativa.

En primera instancia está la relación profesor-alumno. Martínez Navarro (2010) habla sobre las actitudes que requiere tomar el profesor hacia los alumnos con respecto a la evaluación. El autor se expresa en favor de una evaluación continua que mejore el proceso enseñanza-aprendizaje a través de la retro-alimentación para el alumno. En cuanto a la calificación Martínez Navarro (2010) comenta que debe ser justa y equitativa refiriéndose a que la calificación debe mostrar “el nivel de logro de aprendizaje del alumno” y equitativa, es decir aplicando los mismos criterios a todos los alumnos.

Cabra-Torres (2010) considera la evaluación como una relación asimétrica entre el evaluador y el evaluado, en donde el primero es el “experto proveedor de información y conocimiento indiscutible” (Cabra-Torres, 2010:241) y puede emitir juicios sobre el segundo, quien se limita a aceptarlos. En relación con dicha desigualdad, Cabra-Torres enfatiza la cualidad informativa de la evaluación para entender qué sucede en el proceso educativo, o lo que ha denominado “comunicación facilitante”, que repercute en un diálogo entre los participantes del proceso. La propuesta de Cabra-Torres se basa en este diálogo como factor de cambio en el enfoque de la evaluación y que además contribuye a la

democratización del proceso evaluativo, dando un giro a la relación lineal y coercitiva hacia una dialógica y reflexiva que ayuda a crear comunidad entre los participantes del proceso.

Lo anterior es congruente con la enunciación de House (1997:175), quien propone cuatro valores fundamentales para la evaluación educativa: igualdad moral, la autonomía moral, imparcialidad y reciprocidad. Estos valores establecen equidad para todos los individuos; derecho de elección, es decir que nada puede imponérsele a una persona; que no se favorecerá a ningún individuo o grupo; y un sentido de comunidad, que debería radicar en todos los actos humanos.

Moreno Olivos (2016) advierte que la evaluación no puede verse solamente como un aspecto técnico de la educación. El autor señala que el ejercicio de la práctica de evaluación conlleva el ejercicio del poder, el desequilibrio de fuerzas, consecuencias en ocasiones inadvertidas por el docente, y que estas consecuencias son resultado de una supuesta neutralidad, que en realidad se ve afectada por componentes político-ideológicos y ético-morales del profesor y que pueden ocasionar una evaluación sesgada. No es posible, de acuerdo con Moreno Olivos, la objetividad absoluta en la evaluación educativa.

Adicionalmente, sobre la evaluación en la educación, Casero Martínez (2010), quien realiza un estudio cualitativo sobre las opiniones de alumnos acerca de la calidad docente y entre sus hallazgos comenta que el alumno demanda coherencia y justicia en la evaluación, es decir que los contenidos del examen sean coherentes con lo aprendido en clase y los criterios sean justos, sin considerar prejuicios del docente hacia el alumno. Al igual que en el texto de Martínez Navarro, los alumnos en el estudio de Casero Martínez se manifiestan en favor de una evaluación continua. Los alumnos consideraron que los mejores docentes son aquellos que ejercen una autocrítica en su ejercicio y en el momento de la evaluación.

Otro aspecto a considerar que surge del estudio de Casero Martínez es el hecho de la preferencia de los alumnos por los exámenes de desarrollo sobre las pruebas tipo test, las cuales le otorgan mala imagen al profesor debido a su simplicidad en el momento de calificar. Zoeckler (2005) encontró que las calificaciones en una evaluación generalmente se ven influenciadas por factores como esfuerzo, actitud o conducta, los cuales son tomados en cuenta inconscientemente por los profesores llevándolos a emitir juicios y afectar las calificaciones injustamente.

Se considera pertinente culminar con la recomendación de Moreno Olivos (2016:274) al docente de formarse con profundidad en evaluación, lo que lo llevará a “analizar la problemática y complejidad de esta disciplina” y “reconsiderar sus fundamentos teórico-epistemológicos, ético-filosóficos y político ideológicos”.

De esta manera, puede concluirse que la ética del docente en su práctica evaluativa se da en su relación con el alumno, y en la cual deben permear la justicia, la equidad, imparcialidad, neutralidad y reciprocidad en un afán de nivelar dicha relación de naturaleza desigual.

Una vez que se ha realizado un breve recorrido a través de la ética en la educación y principalmente en la labor del profesor universitario, se tocará el tema con respecto al área específica de este estudio, el diseño de exámenes de lengua.

2.3 La ética en el diseño de exámenes de lengua

La ética es un área del conocimiento que no había sido muy explorada en su aplicación a la evaluación de lenguas y, por lo tanto, un área con mayor especialización como lo es el diseño de exámenes de lengua (DEL) es todavía un área de oportunidad de mayor proporción; aun así, el terreno no es inhóspito y varios autores se ocupan ya del tema (Kunnan, 2000, 2004; McNamara, 2000; Shohamy, 2000; Davies, 2004; McNamara y Roeber, 2006; Bachman y Palmer, 2010). Por ejemplo, Hamp-Lyons (2000) desarrolla una discusión sobre la terminología que debe usarse al estudiar la ética del diseñador de exámenes. La autora se inclina por el término ‘responsabilidad’ que implica reciprocidad: “Aceptamos nuestra responsabilidad por los efectos y consecuencias de las actividades de evaluación en las que nos vemos involucrados⁶” (Hamp-Lyons, 2000:582) y vincula el término responsabilidad con lo mencionado por Davies acerca de las tres morales social, profesional e individual.

Farhady (2009) hace una revisión del desarrollo del diseño de exámenes de lengua como área de especialización del lingüista aplicado, definiendo dos perspectivas históricas, la instruccional y la psicométrica; la primera en la que los docentes de lenguas diseñaban los exámenes con base en su propio, conocimiento, gusto, o idiosincrasia, sin tomar en cuenta aspectos como validez o confiabilidad, que fue lo que definió la segunda etapa, la

⁶ Traducción del autor.

psicométrica, que es la que a la fecha ha marcado el ejercicio del DEL. Farhady (2009) describe cómo es en este momento histórico del DEL que se abre el terreno para los exámenes estandarizados, que son diseñados bajo estrictas normas psicométricas, lo que trae a su vez un nuevo problema: al generalizar y estandarizar estos exámenes se dejan de considerar los contenidos de lo que se les enseñó a los alumnos por cumplir con objetivos generales y estandarizados. Estos exámenes estarían poniendo a prueba conocimiento al que los postulantes o candidatos podrían no haber tenido acceso durante su proceso de enseñanza-aprendizaje de lengua, lo cual a su vez no sería justo o equitativo. Derivado de esto, autores como Bachman (2000) y Davies (2004) mencionan que una nueva etapa de desarrollo habría de presentarse en el nuevo milenio, la cual podemos llamar la perspectiva ética, y es que después de la etapa psicométrica y el auge de los exámenes estandarizados, se cuestiona el uso de éstos con diversos fines, distintos a los que fueron diseñados originalmente. Estos autores sugerían mayor investigación en el área, que promoviera la labor del diseñador de exámenes como más justa y equitativa.

Los exámenes de lengua, en general, y no solamente los exámenes estandarizados, han sido cuestionados, especialmente durante los últimos 20 años, ya que debido a la relevancia de las decisiones que se toman con base en ellos se les ha puesto bajo una severa vigilancia. De igual forma, a consecuencia de la importancia de esas decisiones a los diseñadores de exámenes de lengua se les otorga una gran responsabilidad y, junto con ésta, un alto grado de poder; es así que, de acuerdo con Farhady (2009), el diseñador de exámenes debe hacerse responsable de todas las consecuencias que pudiesen derivarse del resultado de su instrumento. De acuerdo con este autor, el diseñador de exámenes de lengua adquiere tres responsabilidades principales: Primero, definir claramente el constructo o características a medir, segundo, definir cómo se va a medir este constructo y; tercero, en considerar las posibles consecuencias que deriven de los resultados de dicha medición. Para Bachman y Palmer (2010) el diseñador de exámenes tiene toda la responsabilidad en otorgar a los interesados completa confianza en que el resultado mostrará una evidencia sólida de lo que están buscando; pero también cuestionan si la responsabilidad del diseñador termina ahí, si se limita exclusivamente al diseño del instrumento.

Davies (2011) enfatiza la necesidad del evaluador de lenguas de desempeñarse conforme a lo establecido en códigos de ética. Este autor, retomando la postura de Bachman (2000),

también promueve la profesionalización del diseñador de exámenes (Davies, 1997, 2011), a través de la literacidad en evaluación (*assessment literacy*) que es definida por Taylor (2009) como lo que profesores e instructores de lengua requieren saber sobre evaluación, es decir el nivel de conocimiento, habilidades y entendimiento de la práctica y principios de evaluación requerido por los involucrados en estos procesos (Taylor, 2009:24). Davies (en Taylor, 2009) enfatizan cómo el campo del diseño de exámenes de lenguas ha evolucionado de ser una labor técnica, a la necesidad de conocimiento sobre lengua y su medición y de aquí al entendimiento de principios como ética y validez. Es así, que Davies (2008) concluye que la profesionalización, lograda a través de la literacidad en evaluación, se compone de habilidades, como manejo de metodología, de diseño de reactivos, estadística y capacidad de análisis de exámenes; experiencia y conocimiento en cuanto a medición, contextualización y descripción de la lengua; y de principios como el uso adecuado de los exámenes, justicia, objetividad, consideración del impacto de los resultados y ética.

Farhady (2009) discute ampliamente la naturaleza de poder de los exámenes de lengua, la cual se debe a que éstos proveen información acerca de los candidatos y esta queda en custodia y poder de los aplicadores. Éstos a su vez proveen la información a los responsables de tomar decisiones; ambos asumen que esta información se ha obtenido a través de medios objetivos válidos y confiables y representan una verdad acerca del candidato, lo cual ejerce una importante influencia en todos los que intervienen en el proceso de evaluación. El poder investido indirectamente en el diseñador de exámenes le exige actuar de manera equitativa, justa, ética, basada en principios morales; en este sentido existe un consenso en el gremio.

No obstante, el diseñador de exámenes, de manera individual, tomará las decisiones de acuerdo con su situación particular, considerando su experiencia, sus valores y su contexto. Hamp-Lyons (2000) lo expresa de la siguiente manera:

Para el evaluador como individuo, los estándares de conducta y códigos de ética tienen gran importancia. La necesidad de discutir en encuentros profesionales cuestiones de responsabilidad, ética, moral, exigencias de profesionalismo, conveniencia de prácticas específicas y decisiones dentro de contextos actuales, sociales, culturales, y políticos, el curso de la 'conducta adecuada' entre los evaluadores, no puede exagerarse. No obstante, al

final cada individuo hará elecciones basándose en su conocimiento, experiencia, valores, limitaciones y prioridades. El diálogo con pares de profesión otorga una ayuda vital para esa toma de decisiones, aunque al final es responsabilidad de cada individuo.” (Hamp-Lyons, 2000).

Considerando lo anterior, con respecto a que las decisiones que toma el diseñador de exámenes de lengua son siempre de manera individual, a pesar de los códigos impuestos por las agrupaciones profesionales o por las instituciones, encontramos una falta de consenso en cuanto a los alcances de la responsabilidad del diseñador de exámenes, ante esto nos encontramos con dos posturas: Primera, el diseñador debe ser responsable, completamente, de las consecuencias que derivan de la aplicación de su examen; segunda, la responsabilidad del diseñador se limita a la elaboración del instrumento, eximiéndole de lo que pueda suceder más allá de la aplicación del test.

En contraste, en la mayoría de los contextos las decisiones sobre cuáles instrumentos deben aplicarse son ajenas al diseñador, así como el uso que se le da a los exámenes o la manera en que se administran o aplican. Por otro lado, cómo podría el diseñador de exámenes realizar instrumentos con la rigurosidad que se requiere de un profesional si no existe una responsabilidad. Por ello es tan importante la ética de trabajo del individuo, que el profesional actúe con base en un código moral, el cual paradójicamente, no debería ser individual sino colegiado.

2.3.1 Enfoques de la ética en el diseño de exámenes de lengua

Los cuestionamientos a los exámenes de lengua se fundamentan justamente en los supuestos anteriores. Un examen de lengua pone a prueba la competencia comunicativa de una persona; sin embargo, es un constructo muy difícil de definir y de medir por lo que la validez de los exámenes de lengua siempre está en tela de juicio. Por otro lado, en varios casos los exámenes de lengua no son utilizados para los fines con que fueron diseñados, como se mencionó renglones arriba, lo que lleva a injusticias o inequidades, hecho que ha sido repetidamente enfatizado por la corriente crítica de diseñadores de exámenes, ya que, si un examen no es justo o equitativo, no puede ser válido (Kunnan, 2000).

Fulcher y Davidson (2000) describen distintos enfoques que existen con respecto al estudio de la ética en el diseño de exámenes. En la primera de estas perspectivas los autores establecen el hecho de que el diseñador de exámenes es parte de una comunidad de profesionales y expertos y existe una responsabilidad hacia estos, y a la vez cada individuo miembro de esta comunidad tiene una responsabilidad con el resto de sus colegas. El siguiente enfoque es acerca del profesionalismo y la democracia. En este se busca el máximo desarrollo del individuo y de la sociedad, y la responsabilidad del examinador de lenguas, sin ser dependiente o independiente de esta comunidad, siempre buscando el bien común. El tercer enfoque es el consecuencialismo y su esencia radica en la consideración de las consecuencias de las decisiones que tome el examinador, o como lo expresan los autores: “La intención de una decisión debería ser el maximizar el bien para la sociedad democrática en que todos vivimos y todos los individuos dentro de esta sociedad⁷” (Fulcher and Davidson, 2005:142). El cuarto enfoque argumenta el uso de los exámenes de lengua como instrumentos de poder. Los autores recopilan una serie de posturas críticas al diseño de exámenes de lengua y documentan el debate en este sentido. La corriente crítica toma impulso desde los pensamientos relativista y postmodernista y cuestiona la práctica del uso de exámenes de lengua con fines políticos, sociales o económicos, de discriminación segregación o jerarquización (Shohamy, 2000; McNamara, 2001).

Una postura más moderada reconoce dicha práctica, pero promueve la reflexión y prácticas basadas en morales y códigos de ética profesionales, así como la rendición de cuentas ante los resultados de los exámenes (Davies, 2004, Kunnan, 2000). Sin embargo, se cuestiona también el hecho de que los exámenes diseñados son, con frecuencia, utilizados para fines distintos al original, situación que escapa del ámbito de competencia del diseñador.

Es así, que la ética en el diseño de exámenes de lengua podría ser estudiada desde dos enfoques que, considerando lo dicho en los párrafos anteriores, están estrechamente relacionados: el primer enfoque es aquel en el que se tiene como principal argumento o punto de interés la validez y, en segundo lugar, el enfoque que tiene como principal énfasis la equidad (*fairness*).

⁷ Traducción del autor.

2.3.1.1 Enfoque en la validez

La validez de las pruebas es un concepto que se desarrolla con la psicometría. En el ámbito de DEL es con Messick (1989, 1990) quien concibe la validez como un concepto aglutinante de varios otros tipos de validez, que hasta ese momento se analizaban de manera separada: validez de contenido, predictiva, concurrente y de constructo. El argumento de Messick sostiene que estos conceptos no necesariamente tienen que ser agrupados, sino que habrá situaciones en que su estudio exigirá que se tengan que separar del resto, sin embargo, cuando hablamos de interpretación de resultados de exámenes, habrá que considerar a la validez como un solo concepto. Messick (1990) sostiene que la validez tiene un efecto en los resultados de las pruebas y estos a su vez afectan la toma de decisiones y ejecución de acciones, con una consecuencia social que no debe ser pasada por alto por el diseñador del test.

Es así que de acuerdo a Messick (1990), en la validación de un examen se debe considerar el significado de la evidencia, y los valores implicados en la interpretación de los resultados y su uso. Esto nos lleva a la consideración de las consecuencias sociales, reales y potenciales, del examen (Messick, 1990:24) y al hecho de que el diseñador del test debe asegurarse que estas consecuencias no sean resultado de un problema de validez en el instrumento. Messick (1990:26) concluye que validez y valores son, ambas, una obligación en el diseño del examen, lo que implica el involucramiento de aspectos científicos y éticos. De aquí la relevancia de lo enunciado por Messick, que en su meticuloso análisis de la validez de los exámenes de lengua muestra el punto de convergencia entre los aspectos éticos del diseñador de exámenes y los técnicos que habían prevalecido hasta ese momento.

2.3.1.2 Enfoque en la equidad (*fairness*)

Kunnan (2000) menciona el concepto de *fairness* como esencial en el estudio de la ética, sin embargo, no concluye con una definición, como puntualizan Bachman y Palmer (2010). Es así que, el término *fairness* se analiza por sus componentes fundamentales, como se mencionó líneas arriba. Siendo uno de ellos el sesgo, el cual McNamara y Roever (2006:82) definen como “una inclinación injusta hacia un lado (grupo, población) en detrimento de otro”.

Dados los diversos aspectos que cubre el término *fairness*, se explica la dificultad de su definición, y a lo que habrá que agregar los problemas en el momento de traducir el término al español. Cuando se habla de *fairness* nos referimos a equidad, imparcialidad, justicia e inclusión. Para fines prácticos, en este trabajo se le denominará en español *equidad*.

El enfoque en la equidad no se separa del enfoque en la validez. La validez es un aspecto inherente a la equidad. Como se mencionó anteriormente; un examen que no es válido no puede ser justo o equitativo (Kunnan, 2000). No obstante, hemos decidido dividir los enfoques con base en la profundidad con que los diversos autores tocan los temas.

En primera instancia habría que definir el término en inglés *fairness*. Kunnan (2000) menciona que no es un término sencillo y que no era muy reconocido en el ámbito de diseño de exámenes. Kunnan no ofrece una definición concreta, sino que busca acercarse a la definición de *fairness* a través de los componentes del concepto. Para Kunnan (2000, 2004) las ideas fundamentales para construir el concepto son validez, ausencia de sesgo y equidad en cuanto a accesibilidad, administración y consecuencias sociales.

Coincidentemente, para Bachman y Palmer (2010) hay dos aspectos esenciales acerca de la equidad y justicia en el proceso de evaluación: Trato equitativo para los individuos y ausencia de sesgo. Así que, considerando el carácter decisivo de la evaluación, Bachman y Palmer (2010) mencionan lo siguiente respecto a la toma de decisiones con base en los exámenes de lengua: “Decisiones justas son aquellas que son igualmente adecuadas, a pesar de la pertenencia de un individuo a un grupo⁸” (Bachman y Palmer, 2010:130) y con respecto a los exámenes de lengua y en este mismo sentido: “Un examen será justo si se les informa los individuos como se tomará una decisión y si esta se toma, en realidad, de la manera en que se le anuncia⁹”. (Bachman y Palmer, 2010:130).

Es responsabilidad del diseñador de exámenes reducir la posibilidad de sesgo, ofrecer a los interesados en los resultados la mayor calidad posible en la información y que la toma de estas decisiones sea, precisamente, justa y equitativa (McNamara y Roever, 2006), siempre buscando que el instrumento sea objeto del menor número de cuestionamientos, para ello,

⁸ Traducción del autor.

⁹ Traducción del autor.

mencionan los autores, es indispensable llevar a cabo un análisis de imparcialidad, al que se refieren como de la siguiente manera: “La revisión de equidad consiste en un proceso sistemático para identificar posibles sesgos en reactivos o reactivos que pudiesen ser suficientemente controversiales como para que la aceptación de un examen se vea dañada¹⁰” (McNamara y Roever, 2006:147), y es que la magnitud de las decisiones puede ser inimaginable cuando se extrapola el problema a ámbitos distintos a la educación.

Autores como McNamara (2000) o Shohamy (2000), por ejemplo, disertan con energía y profundidad en la problemática social en que se ven inmersos los diseñadores de lengua y en el ejercicio del poder por parte de los responsables del diseño de políticas y toma de decisiones. Shohamy (2000) sostiene que el término *fairness* no es exclusivo del ámbito de competencia del diseñador de exámenes, sino que debe estudiarse en todos los contextos adicionales donde las consecuencias de los resultados de estos exámenes tienen alcance; es decir en el uso que se le dé al test.

McNamara y Shohamy (2008) denuncian el uso del poder a través de los exámenes de lengua con fines de nacionalización en distintos países y cómo los diseñadores de lengua deben ajustar su trabajo y decisiones a la voluntad de las personas a cargo de los procesos, y llaman a hacer más conciencia en estos y en crear instrumentos de manera más profesional que disminuyan el grado de inequidad existente en la aplicación de exámenes con estos fines.

Conclusión al capítulo

En este capítulo se ha presentado el aspecto ético en la enseñanza y en la evaluación. Se inició con el establecimiento del acercamiento epistemológico con el que se percibe la ética, que es la complejidad y la ética compleja de Edgar Morin (2006). Enseguida, se realizó la conceptualización de la ética desde una perspectiva general y se acota de manera gradual al ámbito educativo, universitario, de evaluación y de exámenes de lengua. A la vez, se explican algunos conceptos y enfoques relacionados con la ética educativa, como

¹⁰ Traducción del autor.

son alteridad, validez y equidad, esenciales para una visión amplia de la ética en la evaluación.

En el capítulo se exponen como componentes fundamentales de un sistema ético, el profesor y el alumno, quienes participan de maneras egoísta y altruista, incluyente y excluyente, en una especie de vinculación o religación entre sí y con la comunidad y la sociedad. Las acciones de un individuo tomarán uno de estos sentidos, dando lugar a un carácter ético en la toma de decisiones sobre dichas acciones.

Una vez establecida la relevancia del aspecto ético, que se conjugan con los conceptos técnicos relacionados con el tema de evaluación, desarrollados en el capítulo previo, e integran un marco de referencia teórica que contribuye al entendimiento de la temática objeto de este estudio, las prácticas evaluativas de los profesores.

Considerando las bases teóricas en que se basa la mirada a la temática, se expondrán la base epistemológica y metodología de investigación desarrollada para la recopilación de información y sus posteriores análisis.

CAPÍTULO III

METODOLOGÍA DE INVESTIGACIÓN

Introducción al capítulo

En este capítulo se establecen los fundamentos epistemológico, axiológico y ontológico y detallan los procedimientos metodológicos seguidos en el estudio. En primer lugar, se expone la complejidad como un paradigma que permite observar y tratar de entender los fenómenos desde una visión con amplitud, considerando la diversidad de factores que pueden intervenir, así como una declaración sobre los valores y una reflexión acerca del sentido de la realidad y conciencia de existencia de parte del propio investigador. En segunda instancia, se describe y contextualiza el problema, para después explicar los aspectos metodológicos de la investigación como preguntas de investigación, objetivos, enfoques metodológicos, corte y diseño de investigación, y estrategias metodológicas. En seguida, para cada una de las fases de la investigación, se presentan los perfiles de los participantes, los instrumentos de recolección de información y se describen los métodos de recopilación y análisis de la información. El capítulo cierra con una declaración sobre el rigor metodológico aplicado al trabajo de investigación, especialmente en la fase cualitativa.

Cabe adelantar que, para esta investigación se ha adoptado un enfoque mixto, que integra metodología cuantitativa y cualitativa, como se explicará en páginas más adelante. Es preciso aclarar que la presentación de los aspectos metodológicos de la investigación no obedece a su realización cronológica, es decir, primero la fase cuantitativa y después la cualitativa, sino que se busca proveer al lector de una perspectiva amplia, congruente con su carácter mixto, por lo que los elementos del proceso metodológico mencionados en el párrafo anterior se exponen uno a uno, integrando ambas fases.

De igual manera, cabe una aclaración que ha sido resultado del carácter flexible de la investigación de corte mixto. La flexibilidad es una característica de la metodología de investigación en la tradición cualitativa, que permite modificaciones a la investigación a la luz de los hallazgos y que es el caso de este trabajo de investigación. Durante su desarrollo

se plantea el término de responsabilidad profesional, resultado tanto del análisis cuantitativo como del cualitativo, concepto que permite identificar y explicar aspectos de la práctica de evaluación y diseño de exámenes de aprendizaje de lengua de los profesores en el nivel de educación superior. Bajo este entendido, los elementos que conforman la estructura metodológica de la investigación incorporan el concepto de responsabilidad profesional sin haber sido discutido o teorizado previamente. La razón es que es un concepto que fue desarrollado a raíz del entendimiento de la problemática conforme se avanzó en la investigación y que ocupa una sección particular más adelante en el trabajo.

3.1 Fundamentación epistemológica, ontológica y axiológica

3.1.1 La teoría de la complejidad como fundamento epistemológico

El método positivista implica una separación entre el sujeto de investigación y su objeto, una separación entre un conocimiento racional y cualquier otro que no cumple los cánones del método, además de fraccionar el conocimiento y sus disciplinas en especialidades que aíslan el objeto de investigación de otros factores y contextos, en apariencia ajenos, pero que permiten tener una visión más completa de un fenómeno (Borón, 2006).

De acuerdo con Edgar Morin (1990, 2004) la ciencia había tomado un rumbo equivocado, o bien limitado en su esencia, en la generación de conocimiento. La ciencia había buscado describir las leyes bajo las cuales el mundo y el universo se rigen estableciendo un supuesto orden. Bajo los principios de legislar, desunir y reducir, la ciencia había limitado el conocimiento a explicaciones que excluyen todo aquello que no es significativo, que no es susceptible de simplificación y generalización. El objeto de estudio se descompone en partes y se saca de su contexto para enunciar leyes de comportamiento que lleven a un orden generalizado o una predictibilidad explicable y deseable. De esta manera, el conocimiento se rige por una realidad empírica, demostrada y una verdad lógica.

No obstante, explica Morin, esta legislación implicaba que el mundo fuese descompuesto de tal manera que la variedad de factores que intervienen en cada sistema, fenómeno, principio, comportamiento que puede ser sujeto de observación y de estudio pasaba desapercibida, a manera de una ceguera científica, como la llama el sociólogo francés. Riofrío (2001) concuerda en que la ciencia se había vuelto reduccionista y además

asimétrica, lo que implicaba que al explicar un fenómeno se obtenía información sobre otro, resultando en lo que él llama asimetría de la explicación.

En contraste, el pensamiento complejo surgió como una necesidad ante las limitaciones del paradigma científico vigente. Para Morin (1990:32) la complejidad es “el tejido de eventos, acciones, interacciones, retroacciones, determinaciones, azares, que constituyen nuestro mundo fenoménico.” Morin explica que es necesario ver el cosmos como “un proceso en vías de desintegración y, al mismo tiempo, de organización” y no como un sistema o una maquinaria trabajando perfectamente.

El pensamiento complejo ha contribuido al entendimiento y a lograr una mirada desde una perspectiva diferente los fenómenos y eventos en distintas áreas del conocimiento y de manera interdisciplinaria (Prigogine, 1995; Maturana y Varela, 1994; De Almeida, 2008) y de esta manera, han surgido de esta corriente de pensamiento, que buscan el entendimiento y el “conocimiento del conocimiento”, como lo expresa Morin (1990). Por ejemplo, Maturana y Varela (1994) concibieron el concepto de autopoiesis, que consiste en la autorregulación, organización y regeneración de los seres vivos, a manera de sistemas que interactúan entre sí y que se adaptan a los cambios generados por factores externos. Niklas Luhmann desarrolló la teoría general de los sistemas, que establece cómo el orden social moderno se rige a través de sistemas funcionales como son la ciencia, la política, el arte o la religión, con base en la comunicación como un elemento de estos sistemas (Espinosa-Luna, 2013; Bechmann y Stehr, 2002).

De acuerdo con estas teorías de conformación, tanto de los seres vivos como de la sociedad, los sistemas se conforman de individuos autónomos. Tanto los individuos como los sistemas, evolucionan, se reconfiguran y organizan con base en lo que han vivido; generan procesos y relaciones entre sus procesos, que, a su vez se relacionan con los procesos de otros individuos, es decir, se establecen varias relaciones en distintos niveles (Prigogine, 1995). La organización de estos procesos provee identidad a sus organismos y procuran su supervivencia, aunque debido a la interrelación de estos individuos y sus procesos, no es posible distinguir a uno sin su relación con los otros o fuera de su contexto.

La perspectiva de la complejidad aplicada de esta manera a distintos fenómenos permite reconsiderar su legislación, la desunión de sus componentes y la reducción a la

significatividad. Permite una mirada a un todo único que se organiza, reconfigura, permanece, adapta a cada uno de los cambios permanentes en sus procesos, poniendo en duda la suposición del orden y del acatamiento a una ley o generalización postulada con base en una observación limitada a cierto número de componentes del fenómeno.

Según Romero Pérez (2003), la complejidad es una reforma del pensamiento que se origina con base en tres valores epistémicos: saber para hacer, que es combinar conocimientos teóricos con empíricos; saber para innovar, o crear nuevos conocimientos o aplicaciones; y saber para repensar lo pensado, es decir, tomar los conceptos técnicos existentes y explicarlos de una manera comprensible.

Esta corriente de pensamiento también alcanzó la lingüística aplicada y tiene a su mayor exponente en Diane Larsen-Freeman, quien describe la lengua como “un sistema complejo y adaptable que tiene su origen en las interacciones entre los agentes de una comunidad de habla” (Larsen-Freeman, 2011: 49). La autora coincide en que la lengua es un sistema abierto que carece de linealidad, es impredecible, adaptable a nuevas circunstancias, que no es constante: “la lengua es un conjunto dinámico de patrones que emergen de su uso” (Larsen-Freeman, 2011: 52). El carácter social de la lengua, agrega Larsen-Freeman, contribuye a su desarrollo y adaptabilidad, a la construcción de patrones que, cabe señalar, no son resultado únicamente de la instrucción en segunda lengua, sino de toda la experiencia del hablante, incluida su lengua nativa.

Larsen-Freeman puntualiza el reto que representa la investigación en la lingüística aplicada, al volverse imperante no dejar fuera del estudio factores que pudiesen limitar la mirada o el entendimiento a un fenómeno. En este trabajo de investigación se ha tomado la teoría de la complejidad como el paradigma que ayudaría a entender mejor la evaluación del aprendizaje de lengua, los cuales cada uno representa un sistema por sí mismo, la evaluación, el aprendizaje y la lengua.

3.1.2 Fundamentación ontológica

Este trabajo se ha llevado a cabo con una conciencia de parte del investigador de su propio ser y como individuo en una sociedad, a través de las cuales se busca una transformación de vida, tanto para sí mismo, como para otros. Se reconocen las limitaciones con respecto al

conocimiento y capacidades para la consecución de este documento y a la complejidad con que la realidad se construye, siendo esta investigación una aportación limitada a dicha construcción.

Con fundamento en lo anterior, se procedió a la observación del fenómeno en cuestión y al intento de comprender su naturaleza desde su origen, por medio de una interpretación individual y una descripción de un contexto particular, que no escapan a posibles subjetividades o sesgos, resultado del bagaje previo del investigador o de la influencia de la literatura revisada, tal como presuponen Azuaje y González (2018). No obstante, esta investigación se encuentra firme en la búsqueda de propuestas de nuevos marcos de referencia que coadyuven a un cierre de brechas entre dichas subjetividades y a marcar caminos que guíen un acercamiento a la realidad.

3.1.3 Fundamentación axiológica

Este trabajo se ha llevado a cabo con fundamento en los valores asumidos por el investigador, como apertura, respeto, responsabilidad, integridad, imparcialidad, confidencialidad y verdad. La generación de ciencia con fundamento en dichos valores y en las creencias e intereses del investigador tienen como fin desde la perspectiva ética, axiológica y promotora del desarrollo humano, tanto para él como individuo, como para la sociedad de la que forma parte, así como la construcción conjunta del conocimiento, como sugiere Chacón-Arteaga (2014).

Los valores mencionados han regido en la praxis de esta investigación en todo momento, tanto en la interacción con los participantes, así como las acciones de tratamiento, procesamiento e interpretación de la información recopilada de la literatura especializada y de los propios participantes en cada una de las fases.

3.2 Planteamiento del problema

Este trabajo de investigación tuvo inspiración en el llamado de McNamara y Roever (2006) a tomar responsabilidad social sobre los instrumentos de evaluación de lengua. Si bien, el contexto sobre el que se enfocan los autores es diferente al que se toca en este trabajo, ambos coinciden en la necesidad de la búsqueda del aspecto social, humano y a la invitación a reflexionar en la responsabilidad inherente al trabajo de la persona que diseña

los exámenes de lengua, sea en situaciones de alto impacto, como los exámenes de certificación o de impacto menor, como es el caso de los docentes de lengua en el papel de evaluadores en los cursos que ellos mismos imparten. Este trabajo es una contribución a la problemática planteada por McNamara y Roever, y a la vez busca abrir brecha en el contexto mexicano, específicamente en la enseñanza de lengua en el nivel superior de educación.

La evaluación de lenguas tiene como objetivo tomar decisiones con base en la interpretación de los resultados de la información obtenida por medio de instrumentos como exámenes. Los individuos que intervienen en este proceso, administradores, profesores, diseñadores de exámenes, estudiantes, candidatos y usuarios externos, se ven afectados por esas decisiones en distinto grado ya que las consecuencias son de diversa índole, siempre dependiendo de cada caso en particular. Los administradores del proceso deben optimizar los recursos; los profesores y diseñadores de exámenes están a cargo del desarrollo y, a menudo, de la aplicación y calificación del instrumento; los estudiantes o candidatos son los más vulnerables ya que deben seguir los planteamientos y acatar las decisiones de los agentes mencionados y generalmente no son consultados sobre el proceso y mucho menos intervienen en las decisiones; y los usuarios externos que son aquellos que hacen uso de la información, resultado del proceso de evaluación pero que no intervienen directamente en él. El proceso de evaluación debe estar planeado y coordinado entre todos sus integrantes para que las decisiones se tomen de manera informada, fundamentadas en interpretaciones significativas acerca de los resultados obtenidos.

Con miras a lo anterior, la psicometría y la lingüística aplicada se han encargado de estudiar, elaborar y mejorar los instrumentos de evaluación de lenguas, de acuerdo con McNamara y Roever (2006); sin embargo, aún está pendiente el aspecto social, que en estos estudios ha pasado a un segundo término ya que desarrollar instrumentos que midan los constructos con mayor precisión se había convertido en una prioridad para los investigadores del área. Los autores agregan que un instrumento psicométricamente bien realizado, no necesariamente es socialmente adecuado (McNamara y Roever, 2006) y las consecuencias a nivel social y personal pueden ser de la mayor importancia. Es en este sentido McNamara y Roever (2006), al igual que McNamara (2001), detectan un vacío en

la lingüística aplicada, en cuanto a la investigación sobre el aspecto social de la evaluación de lenguas se refiere.

Otros autores también hacen un llamado a la necesidad de replantear o reenfocar la cultura de la examinación (Spolsky, 2008, Bordón 2015). Stansfield (2008) hace notar la creciente necesidad de examinadores de lenguas, pero a la vez llama a una profesionalización y concientización de la responsabilidad de éstos. En este mismo sentido, Davies (2008), Taylor (2009) y Fulcher (2013) trabajan con el término *assessment literacy* para referirse a la necesaria profesionalización del diseñador de exámenes.

Al realizar una revisión de la literatura mencionada sobre diseño de exámenes de lengua se han encontrado una gran cantidad de textos destinados a la preparación técnica y diseño de exámenes, a las necesidades de literacidad en evaluación, a los procedimientos ideales o recomendados por expertos, a hacer una crítica sobre el papel de los evaluadores y diseñadores en los exámenes de alto impacto y su responsabilidad social, así como una gran cantidad de estudios cuantitativos sobre técnicas y exámenes en distintos contextos.

Sin embargo, poco se ha indagado y escrito sobre el diseño de exámenes en México, sobre las prácticas evaluativas reales del docente, sobre sus creencias, motivaciones, razones y justificaciones, que junto con el contexto en que se insertan, nos permitan entender la evaluación, y en particular la práctica de diseño de exámenes de los profesores de lengua extranjera en educación a nivel superior. Una aportación de esta investigación va en este sentido, en la indagación y documentación de dicha práctica, además de identificar en ella rasgos de literacidad en evaluación y en los aspectos ético, social humanístico y profesional del docente. Todos estos factores, así como los antecedentes de cada docente en particular generan una problemática que se entreteje de manera compleja, razón por la cual se ha decidido explorarla, justamente, desde la perspectiva del pensamiento complejo (Morin, 1990).

Este trabajo también busca contribuir en este sentido, al aplicar un fundamento epistemológico que busca construir conocimiento desde el entendimiento integral de un fenómeno multifactorial, que es la evaluación del aprendizaje de lenguas y que generalmente es estudiado con visiones analíticas que no logran una percepción completa de la problemática. Es así que, el diseño de exámenes es estudiado con una percepción del

individuo, en este caso el docente, con antecedentes, ideologías, prácticas diferentes, en distintos grados, a las de sus colegas, que pueden ser irregulares, falibles o inconsistentes, pero que a su vez pueden mostrar profesionalismo, responsabilidad y ética de trabajo, que en conjunto nos abran una ventana a la realidad del profesional de docencia de lenguas y en particular de su evaluación.

Como se ha mencionado previamente, este estudio se llevó a cabo en una universidad estatal en el Estado de México, particularmente en el campus principal, en la ciudad de Toluca. Esto debido al interés del investigador por conocer a profundidad la problemática descrita en este inciso en su contexto cercano e inmediato, lo cual permitiría, a su vez, tener un pronto impacto en la práctica real, tanto propia, como de sus colegas y la institución. Los sujetos de investigación son los profesores de segunda lengua, inglesa o francesa, en los estudios a nivel superior en la UAEMex. Estos profesores trabajan en una o varias facultades y son responsables de la elaboración de los exámenes para sus grupos, ya sea de manera individual o colegiada. A pesar de ser el contexto de trabajo del investigador y de conocer algunas de las prácticas evaluativas del profesorado, no existía una certeza o conocimiento sobre los fundamentos o razones para dichas prácticas, como tampoco existía documentación o antecedentes sobre la temática en este contexto en particular. En seguida se presentan las bases metodológicas del estudio que busca lo anterior, la descripción de la práctica evaluativa y las razones detrás de ella.

3.3 Preguntas de investigación

La investigación dio inicio con una pregunta que planteaba identificar las prácticas evaluativas del profesor de lengua en su papel como diseñador de exámenes. Sin embargo, este cuestionamiento dio pauta a nuevas preguntas que se pensó debían ser exploradas para entender mejor la pregunta inicial. Por consiguiente, las preguntas que finalmente constituyeron la estructura y metodología de esta investigación fueron las siguientes:

- ¿Cuáles son las prácticas del profesor de lengua como diseñador de exámenes en la evaluación del aprendizaje de lenguas?
- ¿Cuáles son los factores que sirven como base para la toma de decisiones en la práctica evaluativa por parte del profesor?

- ¿Cómo construye el profesor de lengua-diseñador de exámenes su responsabilidad profesional a través de su discurso?
- ¿Cómo se interrelacionan las dimensiones de la responsabilidad profesional y sus elementos en el discurso del profesor de lengua-diseñador de exámenes?
- ¿Cómo se manifiesta la literacidad en evaluación en el discurso de los profesores de lengua?

3.4 Objetivos

Para contestar estas preguntas se plantearon los siguientes objetivos:

3.4.1 Objetivo general

- Deconstruir la práctica evaluativa del profesor de lengua como diseñador de exámenes de lengua extranjera.

3.4.2 Objetivos particulares

Al igual que las preguntas de investigación y con el fin de lograr coherencia con éstas, los objetivos de la investigación se modificaron obedeciendo a los hallazgos y a las acciones para contestar dichas preguntas, lo que resultó en los siguientes objetivos planteados:

- Examinar las prácticas del profesor de lengua en diseño de exámenes de lengua extranjera.
- Determinar los factores de la práctica evaluativa en que se basa el profesor para la toma de decisiones.
- Identificar de qué manera el discurso del profesor de lengua da cuenta de su práctica en diseño de exámenes y responsabilidad profesional.
- Desarrollar y aplicar instrumentos de corte cualitativo y cuantitativo que permitan identificar las prácticas del profesor de lengua en el diseño de exámenes de lengua.

3.5 Enfoque metodológico

La metodología mixta es aquella que integra elementos de los enfoques cuantitativo y cualitativo en un mismo estudio o en una serie de investigaciones (Hernández Sampieri, 2008; Brown, 2014, Creswell, 2014, Mertens 2014); es decir, que las preguntas de investigación, los procedimientos de recolección de datos, los tipos de análisis o los tipos de conclusiones de ambos enfoques sean utilizados por el investigador para estudiar un fenómeno desde las dos perspectivas (Tashakkori y Creswell, 2007). Esto le permite al investigador comprender un fenómeno desde diferentes ángulos que serían imposibles de visualizar desde una sola perspectiva, además de que al triangular la información se valida y refuerza una a la otra (Brown, 2014).

De acuerdo con Larsen-Freeman (2011), el estudio de fenómenos desde la perspectiva de la complejidad no puede realizarse a través del enfoque cuantitativo. Este enfoque positivista, necesariamente requiere la descomposición en partes y aislamiento de los factores, por lo que el enfoque cualitativo permite un entendimiento más completo de la problemática, al contemplarse todos los factores que intervienen.

En este trabajo se ha decidido adoptar un enfoque mixto para recopilar información y poder entender la problemática de manera integral. No obstante, la fase cuantitativa se utilizó para dar validez a la fase cualitativa, y tuvo un alcance descriptivo que permitió una visión general de la problemática que pudiera dar lugar a aspectos de relevancia en la fase cualitativa.

A continuación, se describen las etapas de recopilación de información, así como el corte y diseño de investigación.

3.5.1 Corte y diseño de investigación

3.5.1.1 Corte de investigación

Esta investigación se llevó a cabo en dos etapas, la primera, cuantitativa, en la que se aplicó una encuesta entre los participantes y la segunda cualitativa, en la que se realizaron entrevistas semiestructuradas por lo que el estudio es de corte mixto. En la siguiente tabla se muestran las etapas de la investigación:

Tabla 3.1*Etapas de investigación*

Etapa	Enfoque utilizado	Instrumento utilizado para la recolección de datos	Tipo de análisis de datos	Fecha
1	Cuantitativa	Encuesta	Análisis factorial exploratorio	2016
2	Cualitativa	Entrevistas	Teoría fundamentada Análisis del discurso	2017

El uso del enfoque mixto es el caso de esta investigación en la que en la primera fase se aplica a una muestra del universo, el cuestionario cuantitativo que permitió identificar las prácticas en cuanto a diseño de exámenes y evaluación de lenguas de los docentes de una institución de educación superior. Para explorar y conocer sobre estas prácticas fue necesario levantar una encuesta debido al tamaño del universo bajo estudio ya que este instrumento nos permitió explorar una generalidad y detectar aquellas prácticas que prevalecen.

Sin embargo, en la etapa posterior el enfoque cambió en favor del corte cualitativo. Esta decisión se dio debido a que, para conocer la práctica evaluativa de manera más profunda y realmente entender las razones por las que los docentes-diseñadores de exámenes toman ciertas decisiones, se pensó que el análisis del discurso sería una herramienta más adecuada, que permitiría explorar el pensamiento del individuo, en lugar de una práctica colectiva. De igual manera, el análisis del discurso permite un acercamiento más preciso y detallado de las enunciaciones de los profesores en torno a la práctica evaluativa y la responsabilidad profesional.

La relación entre la educación y el discurso es sumamente estrecha como lo explica Buenfil Burgos (1993), quien hace una aproximación entre lo discursivo y lo educativo. Conceptualiza el uso académico del discurso como una declaración de principios y objetivos y comenta el carácter de lo educativo que es relacional, abierto y precario e inestable (Buenfil Burgos, 1993). Este carácter de lo educativo implica que: en primera instancia, no es posible

definir la educación separada de un discurso; segundo, que no es posible identificar todas las características, elementos y prácticas que conforman la educación; y tercero que el proceso educativo no es estable y sucumbe ante factores externos. Consecuentemente, dichos principios y objetivos en la práctica educativa se construyen dentro de un discurso, con un sinnúmero de factores y componentes y que además carecen de una estabilidad debido a factores externos. Dado lo anterior, surge la imposibilidad de comprender la práctica educativa mediante conclusiones generalizadas por medio de métodos cuantitativos, sino que es por medio del discurso que puede hacerse ese acercamiento a su contexto, identificar componentes, así como los factores internos y externos, y así lograr comprensión, o al menos un acercamiento a la comprensión del individuo y su relación a lo educativo.

3.5.1.2 Diseños de investigación

Al ser una investigación mixta, en este trabajo se combinan diferentes tipos de diseño para las diferentes fases, la cuantitativa y la cualitativa.

En primera instancia, para la fase cuantitativa, en la cual se aplicó una encuesta y se seleccionó un diseño de tipo no experimental, transversal y descriptivo. El diseño de esta fase es no experimental, ya que se trata de hacer una descripción sin manipulación de variables. Se busca identificar las prácticas evaluativas y de diseño de exámenes en su contexto propio, tal y como son, por lo cual es descriptivo; y en un momento en particular, de ahí que el diseño sea transversal.

Para conocer cómo se estructura la práctica evaluativa del profesor en su discurso, se consideró necesario, en primera instancia, conocer la práctica docente en evaluación, ubicarla en un contexto en un momento y espacio, a lo que se debe la transversalidad del estudio y su carácter descriptivo. El conocer las prácticas en evaluación y diseño de exámenes de los docentes nos ha permitido conocer el contexto de manera general. La fase cuantitativa nos dio la oportunidad, a través del cuestionario, de tener una idea muy clara de la labor del docente universitario en evaluación de lenguas, se logró observar un comportamiento general que dibuja un perfil, una construcción de la práctica evaluativa y de diseño de exámenes de lengua. Una vez que se logró tener esa visión general del docente, se tomó como base para explorar el perfil a nivel individual.

Dadas las características de la investigación cualitativa es posible ubicar la fase segunda dentro de este enfoque, ya que se utilizaron técnicas de recopilación y análisis de información interactivas y humanistas como las entrevistas y el análisis del discurso; las preguntas de las entrevistas y el tipo de información se adecuaron a la información obtenida en la fase exploratoria; la información se analizó de manera interpretativa; y el estudio se llevó a cabo buscando incluir todos los aspectos que influyen en la práctica evaluativa del profesor de lengua.

Con base en las decisiones metodológicas expuestas en los párrafos previos, puede afirmarse que el diseño original fue secuencial explicativo, enfoque de dos fases en que se realizan una después de la otra (Creswell y Plano Clark, 2018) y donde se busca que la primera fase, cuantitativa, arroje datos que puedan explicarse en una segunda fase, cualitativa. Pereira (2011) menciona, que con base en los objetivos de la investigación e interés del investigador, el estudio secuencial puede otorgar un “estatus dominante” a una de las fases, como es el caso de este estudio, donde la fase cualitativa tomó relevancia una vez que la información se organizó, analizó e interpretó, además del planteamiento epistemológico, donde la teoría de la complejidad armoniza con el enfoque cualitativo. De este modo, los datos cuantitativos fueron utilizados para: En primer lugar, explorar el estado de la práctica evaluativa de la muestra en general; en segundo lugar, dar validez al análisis cualitativo.

De acuerdo con lo mencionado en párrafos anteriores, con respecto a la adecuación del procedimiento y análisis a la naturaleza de la información obtenida en el estudio exploratorio y en las entrevistas, también puede afirmarse que el estudio es de diseño flexible. Maxwell (En Mendizábal, 2006) define éste como la “disposición de elementos que gobiernan el funcionamiento de un estudio”. Mendizábal (2006) explica que estos elementos son sujetos de modificaciones o articulaciones que permitirán la producción de datos cualitativos de manera inductiva. Es así como, esta característica del diseño de investigación cualitativa ha permitido al estudio tomar cursos de acción conforme a la naturaleza de la información obtenida. Se planteó identificar las prácticas evaluativas de los sujetos en la fase cuantitativa. Una vez levantados los datos acerca de las prácticas evaluativas, fue posible indagar a profundidad en ellas mediante las entrevistas, en las que

los profesores se manifestaron acerca de sus perspectivas sobre el diseño de exámenes y evaluación de lengua.

3.5.2 Estrategia metodológica

3.5.2.1 Estudio de caso

Dentro de las tradiciones metodológicas de investigación cualitativa (Creswell, 2014) se encuentra el estudio de caso. Yin (2003) lo define como el estudio de un fenómeno dentro de su contexto real, en especial cuando los límites entre ellos no están claramente definidos. Para Yin el contexto es trascendente para el estudio y debe analizarse con profundidad. Creswell (2014) lo define como la exploración de un sistema delimitado en cuanto a tiempo y lugar, y puede referirse a un caso particular o casos múltiples que son estudiados a profundidad dentro de su contexto. El objeto de estudiar un caso es entender el fenómeno con claridad y sin pretender hacer generalizaciones (Stake, 1995), ya que dada la especificidad del contexto la información obtenida serviría exclusivamente para entender el fenómeno en cuestión y poco probable que pueda traspolarse a otro contexto de manera natural. Como se mencionó líneas arriba, es preciso realizar el estudio dentro del contexto natural. Transferir las condiciones dadas a un contexto diferente significaría información distinta y por ende un caso distinto.

El caso que nos ocupa es el de los profesores de lengua inglesa y francesa de la UAEMex que son a la vez diseñadores de exámenes de lengua en programas de educación superior. Consideramos que este estudio es pertinente, especialmente porque se incluye a profesores de la Licenciatura en Lenguas. Una característica distintiva es que los estudiantes de esta licenciatura al egresar se convierten en docentes de lengua inglesa y francesa por lo que lo que ellos observan en sus clases de lengua tiene una repercusión en su práctica profesional, es por esto por lo que el trabajo del docente de lengua en este programa de licenciatura es de alto impacto y relevancia.

3.5.2.2 Teoría fundamentada

Para el análisis de la información recopilada en las entrevistas en esta fase se adoptó el paradigma cualitativo, con el método de Teoría Fundamentada, que fue desarrollada por Glaser y Strauss (1967) y más adelante por Strauss y Corbin (1990, 1998). Este método

consiste en la recopilación, análisis y codificación de los datos de manera simultánea (Bonilla García y López Suárez, 2016), que permite el desarrollo de teoría con base en los mismos datos. La metodología de la Teoría Fundamentada precisa tres procesos:

- Codificación Abierta
- Codificación axial
- Codificación selectiva

La codificación abierta consiste en separar los datos en unidades de información y categorizarlas (Moghaddam, 2006). De acuerdo con San Martín (2014), mediante una minuciosa revisión de un texto, la codificación abierta resulta en la revelación de conceptos, ideas, sentidos, pensamientos y significados contenidos en él. En un momento del análisis, el investigador se encontrará con lo que se ha llamado saturación teórica (Moghaddam, 2006; San Martín, 2014), fenómeno que se da en el momento en que no surgen nuevos códigos y que las relaciones entre ellos se han establecido. La saturación teórica permite al investigador llegar a una conclusión en el análisis.

Una vez que se ha iniciado la codificación abierta, se da también el siguiente proceso de la teoría fundamentada. La codificación axial consiste en relacionar las categorías que surgen en la codificación abierta, de manera que el número de códigos derivados de la codificación abierta se reduzca en número (Moghaddam, 2006). Esta síntesis de códigos resulta en el siguiente proceso que es la codificación selectiva, donde los códigos del proceso anterior convergen en un código central que representa la teoría y que pudiera explicar el comportamiento que se está estudiando.

Dicha convergencia se da por el establecimiento de relaciones entre los códigos y las categorías surgidas de los constantes análisis y comparaciones de parte del investigador. Este proceder es denominado Método de comparación constante, que de acuerdo con San Martín (2014), consiste en comparar constante y permanentemente los códigos que surjan del análisis con los que se hayan originado previamente. Esta comparación permite al investigador establecer hipótesis sobre las relaciones entre los códigos, establecer causas, consecuencias, jerarquías, procesos, tipos o condiciones. No obstante, la teoría fundamentada no busca la comprobación de las hipótesis que se originen del estudio sino el establecimiento

de la teoría resultante de las relaciones encontradas entre los códigos y categorías determinadas mediante el análisis de la información.

La convergencia de las categorías en las relaciones entre los códigos y que surgen de las codificaciones abierta y axial resultan en la formulación de hipótesis y conceptualización de términos que mediante la teorización, contribuyan a explicar los datos recopilados. La aportación de la teoría fundamentada se da en este momento de la investigación, que es la concreción del proceso de análisis de la información recopilada.

3.6 Participantes

3.6.1 Sujetos en fase cuantitativa.

En la etapa exploratoria, primera del proyecto, se aplicó el cuestionario a 88 profesores de lengua inglesa y francesa. Este instrumento sirvió para obtener las variables atributivas o personales de los participantes en el estudio, las cuales mostraron la siguiente información.

3.6.1.1 Edad

La mayoría de los profesores se ubica en el rango de los 31 a los 35 años, 38 participantes que representan el 43.2%. Solamente dos profesores (2.3% del total) se encuentran en el rango de más jóvenes, entre 20 y 25 años y solo un profesor rebasa los 50 años, y el resto de manera similar entre los 26 y 30 y los rangos de 36-40 y 41-49, como se muestra en el cuadro y gráfica siguientes:

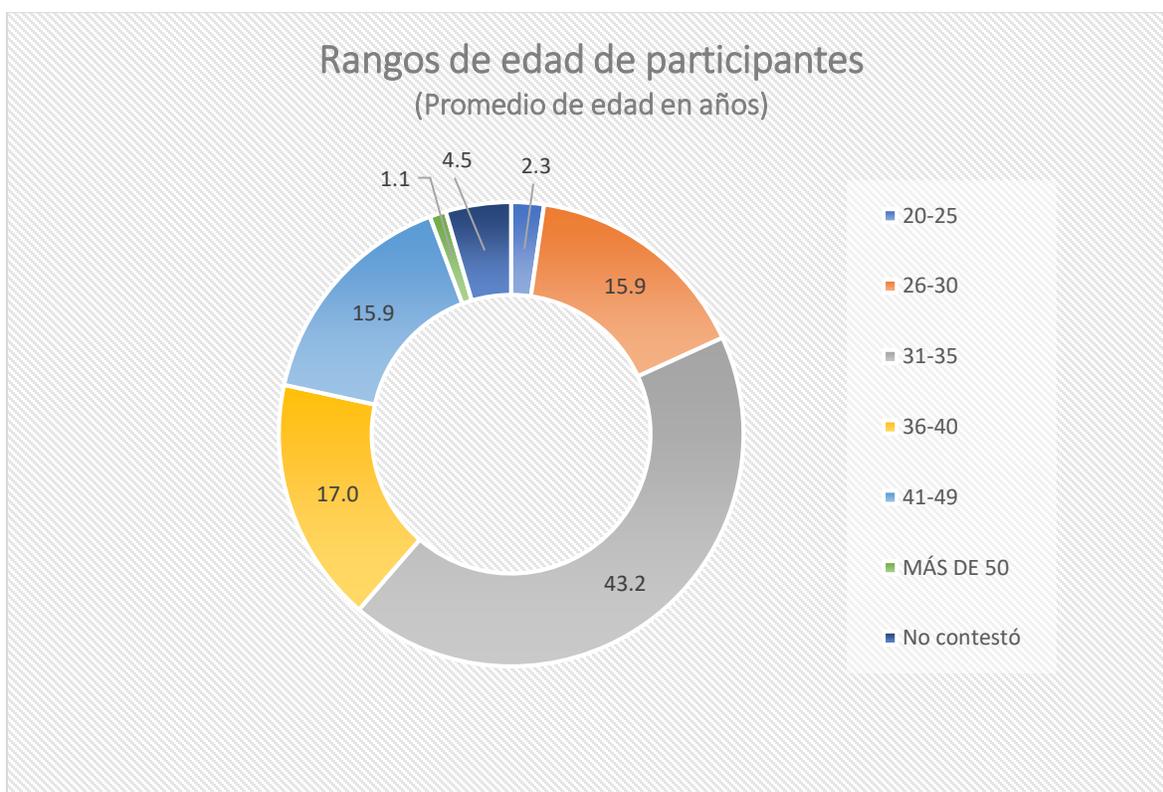
Tabla 3.2

Rango de edad de participantes

RANGO DE EDAD	n	%
20-25	2	2.3
26-30	14	15.9
31-35	38	43.2
36-40	15	17.0
41-49	14	15.9
MÁS DE 50	1	1.1
NO CONTESTÓ	4	4.0
Total	88	100

Gráfica 3.1

Rango de edad de participantes



Nota: Elaboración propia.

3.6.1.2 Género

La gran mayoría de los profesores encuestados son mujeres, 68 que representan el 77.3% y solamente el 20.5%, 18 profesores varones.

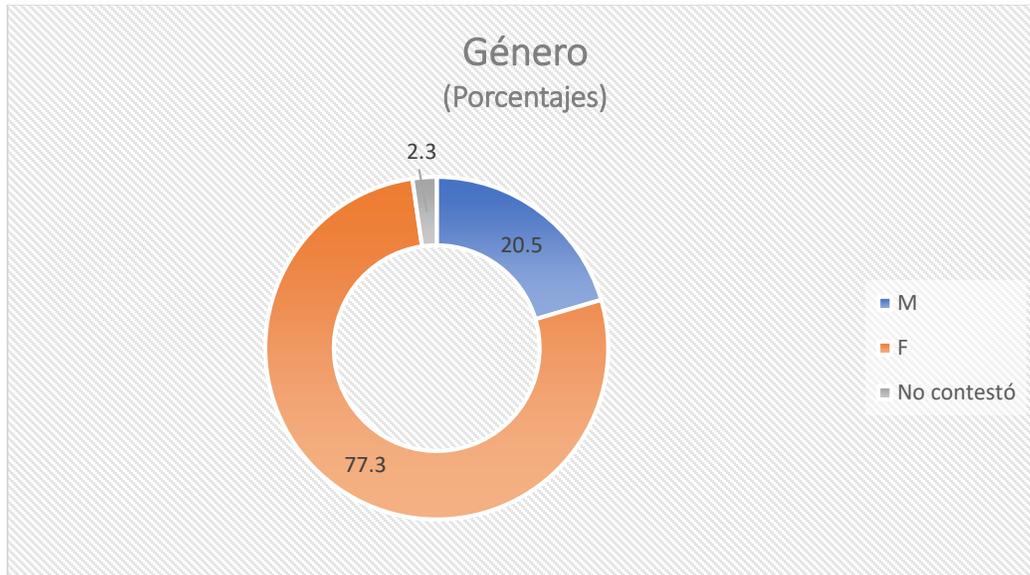
Tabla 3.3

Género de los participantes

GÉNERO	n	%
MASCULINO	18	20.5
FEMENINO	68	77.3
NO CONTESTÓ	2	2.3
TOTAL	88	100.0

Gráfica 3.2

Género de los participantes.



Nota: Elaboración propia.

3.6.1.3 Años de experiencia docente

El rango de años de experiencia va desde 1 a más de 15, siendo en general una plantilla de profesores experimentada, ya que solo el 17% de profesores tiene experiencia menor a los cinco años. Las edades ubican a los profesores a lo largo de los diferentes rangos en números similares, como se aprecia en la tabla y gráfica siguientes:

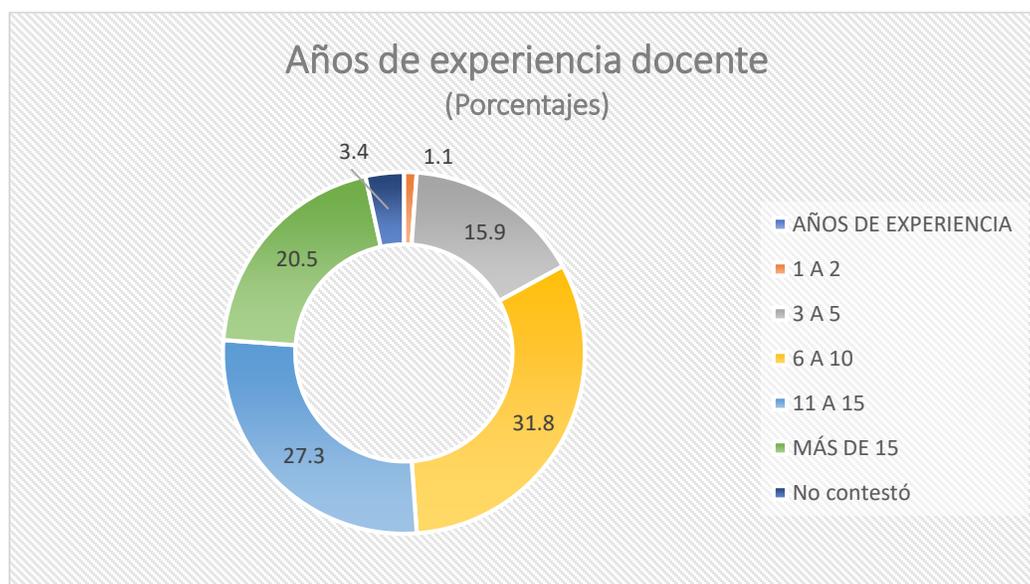
Tabla 3.4

Años de experiencia en docencia de lenguas de los participantes

AÑOS DE EXPERIENCIA	n	%
1 A 2	1	1.1
3 A 5	14	15.9
6 A 10	28	31.8
11 A 15	24	27.3
MÁS DE 15	18	20.5
NO CONTESTÓ	3	3.4
Total	88	100.0

Gráfica 3.3

Años de experiencia en docencia de lenguas de los participantes.



Nota: Elaboración propia

3.6.1.4 Grado de estudios

La mayoría de los profesores de lengua cuenta con un posgrado (53.4%) y 44.3% cuenta con grado de licenciatura, requisito mínimo para ser contratado para impartir clases a nivel de educación superior.

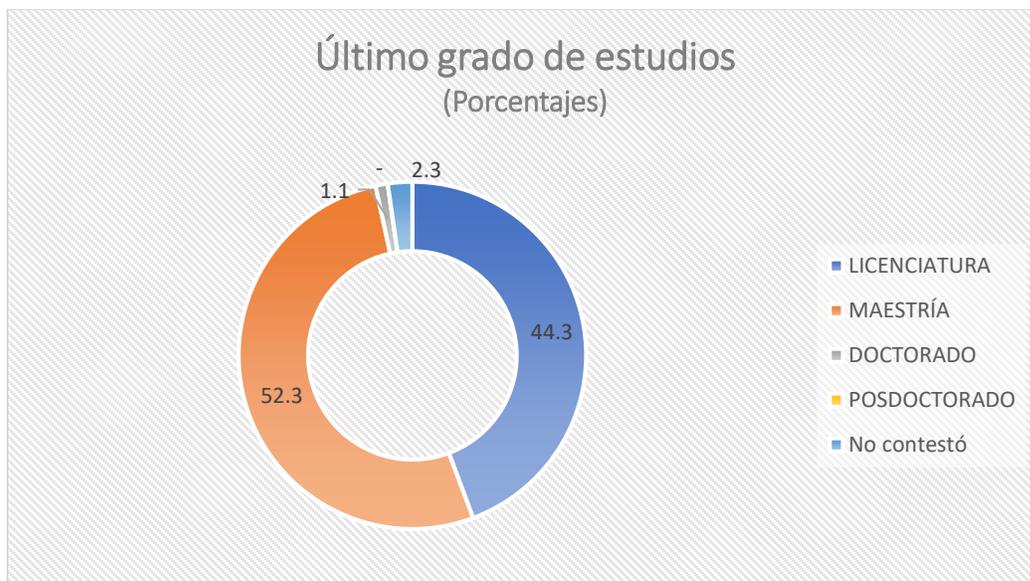
Tabla 3.5

Escolaridad de los participantes

ÚLTIMO GRADO DE ESTUDIOS	n	%
Licenciatura	39	44.3
Maestría	46	52.3
Doctorado	1	1.1
No contestó	2	2.3
Total	88	100.0

Gráfica 3.4

Escolaridad de los participantes.



Nota: Elaboración propia.

3.6.1.5 Sujetos por organismo académico

Por organismo académico nos referimos al lugar de adscripción o de trabajo, de los profesores; cada organismo académico es una Facultad de la UAEMex, y aunque en cuanto a sus planes de estudios todas se rigen por lo que dicta la Dirección de Aprendizaje de Lenguas, con excepción de la Facultad de Lenguas, en cada Facultad los maestros se organizan de manera independiente, en cuanto a la evaluación, por ejemplo, los maestros deciden como debe evaluarse y son responsables por los exámenes al interior de su organismo en particular.

Por su naturaleza, la Facultad de Lenguas es el organismo que más profesores de lengua tiene y la que más aporta a este estudio (37.5% del total de los participantes en el cuestionario). A pesar de que se solicitó la participación de todos los organismos académicos solo algunos accedieron a participar y, aun así, no fue posible contactar a algunos profesores de los organismos que aceptaron colaborar en el estudio o simplemente no respondieron las invitaciones de manera particular. De los 20 organismos de la Universidad, solo fue posible conseguir la colaboración de 13 de ellos. En cuanto al número

de profesores, la Universidad reportó 174 profesores de lengua, que junto con los de la Facultad de Lenguas hace un total de 208 profesores, de los cuales 88 contestaron el cuestionario; esto representa un 42.3 % del total de profesores de lengua de la Universidad.

Se distingue la participación de los profesores de la Facultad de Lenguas, quienes aportan un 37.5 % del total, después los profesores de la Facultad de Arquitectura y diseño (14 profesores, 15.9% del total). La siguiente tabla y gráfica nos muestran cómo se integran los profesores de acuerdo con su organismo académico de adscripción:

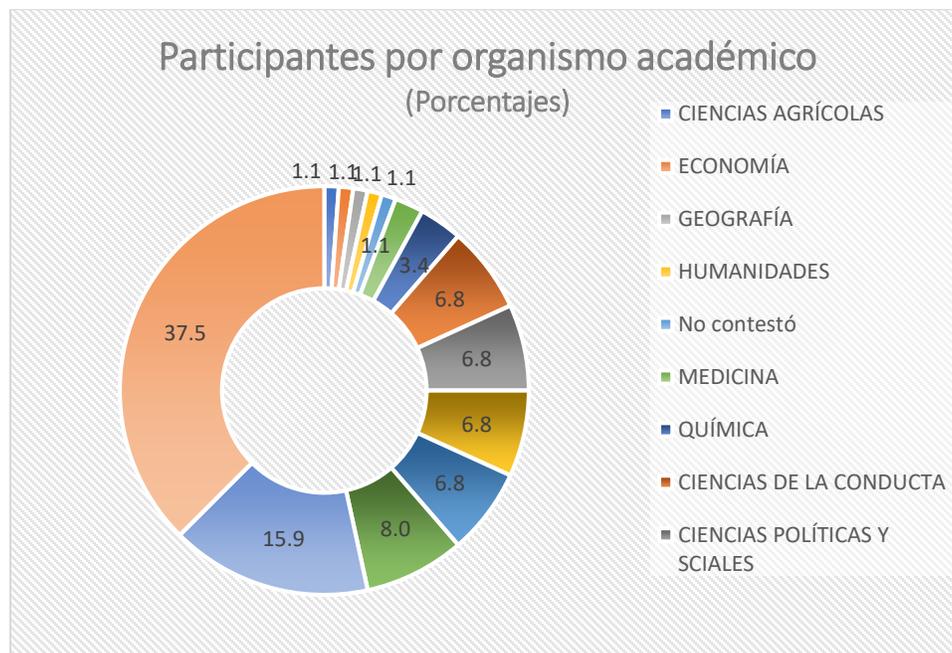
Tabla 3.6

Participantes por organismo académico de adscripción.

ORGANISMO ACADÉMICO	n	%
CIENCIAS AGRÍCOLAS	1	1.1
ECONOMÍA	1	1.1
GEOGRAFÍA	1	1.1
HUMANIDADES	1	1.1
NO CONTESTÓ	1	1.1
MEDICINA	2	2.3
QUÍMICA	3	3.4
CIENCIAS DE LA CONDUCTA	6	6.8
CIENCIAS POLÍTICAS Y SOCIALES	6	6.8
CONTADURÍA Y ADMINISTRACIÓN	6	6.8
DERECHO	6	6.8
INGENIERIA	7	8.0
ARQUITECTURA Y DISEÑO	14	15.9
LENGUAS	33	37.5
Total	88	100.0

Gráfica 3.5

Participantes por organismo académico de adscripción.



3.6.2. Participantes en la entrevista

Al finalizar la encuesta de la fase cuantitativa, los sujetos se encontraron con un apartado donde se preguntaba si estarían dispuestos a participar en una entrevista y encontraban también un espacio donde podían escribir sus datos de contacto en caso de acceder a ser entrevistados. Se procedió a establecer contacto con los participantes en esta situación. Se logró el contacto y las condiciones para realizar las entrevistas con 12 profesores. Los participantes en la entrevista son descritos en la siguiente tabla, que muestra el nombre, en este caso un pseudónimo, para asegurar el anonimato del participante, su edad, grado académico, experiencia en la docencia y la institución donde laboran, toda esta información verídica.

Tabla 3.7*Participantes en las entrevistas.*

	Nombre (Pseudónimo)	Edad	Grado académico	Experiencia docente (Años)	Institución	Lengua
1	Lorena	39	Maestría	18	Lenguas y Derecho	Inglés
2	Erika	26	Licenciatura	5	Ingeniería	Inglés
3	Nora	33	Maestría	10	Lenguas	Inglés
4	Camila	37	Licenciatura	13	Lenguas	Inglés
5	Gina	35	Maestría	9	Lenguas	Francés
6	Alondra	42	Licenciatura	15	Medicina e Ingeniería	Inglés
7	Simón	27	Licenciatura	5	Lenguas y Derecho	Inglés
8	Graciela	26	Licenciatura	3	Lenguas	Inglés
9	Belinda	28	Maestría	9	Lenguas	Francés e inglés
10	Genaro	30	Maestría	8	Lenguas y Humanidades	Inglés
11	Virginia	35	Licenciatura	11	Lenguas	Francés
12	Nadia	31	Licenciatura	7	Lenguas	Francés

3.7 Instrumentos de recolección de información

3.7.1 Instrumento de recolección de información cuantitativa: Encuesta

Con el fin de conocer las prácticas evaluativas y de diseño de exámenes de los profesores de lenguas y poder tener un primer acercamiento, de manera general, a cómo se conforman estas prácticas, se decidió levantar una encuesta. Para ello, se llevaron a cabo una serie de entrevistas informales con profesores de lengua experimentados, cinco en total, a quienes se les pidió enunciar prácticas que hubiesen observado en los diseñadores de exámenes de lengua y así se obtuvieron elementos empíricos para la redacción de reactivos.

Los profesores seleccionados para hacer estas observaciones fueron todos docentes de lengua, con experiencia de entre 18 y 25 años en enseñanza de lengua inglesa y una de ellas en lengua francesa. Además, todos son a su vez, formadores de profesores de lengua y tres de ellos con experiencia en enseñanza e investigación en evaluación de aprendizaje de lenguas. Se consideró que la experiencia de estos profesores podría aportar una visión muy amplia y crítica sobre la práctica evaluativa en la enseñanza de lenguas. En específico se les pidió mencionar acciones de las que tuviesen conocimiento independientemente de la opinión, favorable o desfavorable, acerca de las prácticas que se mencionaran.

De esta manera, se obtuvo un listado el cual se depuró para obtener una serie de enunciados que describían las prácticas de los docentes en cuanto a diseño de exámenes de lengua. Los enunciados se modificaron en su redacción de manera que se plantearan en una oración afirmativa. Algunos ejemplos de lo obtenido después de estas entrevistas informales son:

- a. Los exámenes son calificados únicamente por una persona.
- b. Los exámenes tienen errores ortográficos, de redacción o léxicos.
- c. Los exámenes se pilotean antes de ser aplicados.

Por otro lado, los elementos teóricos y técnicos se obtuvieron después de revisar la literatura referente a las prácticas deseables del docente como profesional de evaluación y diseño de exámenes (Messick, 1989; Bachman y Palmer, 1996; House, 1997; Román, 2003; Mauri, 2003; Martínez Navarro, 2010; Casero Martínez, 2010; Farhady, 2009; Fulcher y Davidson, 2007; Kunnan, 2004; McNamara y Roever, 2006). Una vez revisada la literatura y realizadas las entrevistas informales con los expertos, se conjuntaron las ideas para elaborar los enunciados que describirían el mayor número de prácticas representativas de la labor del profesor-diseñador de exámenes. Estos enunciados pasaron por un proceso de depuración, en la que se aplicaron criterios como evitar repetición, ideas confusas o abstractas y el uso de términos técnicos. El resultado de esta depuración fue un cuestionario de setenta enunciados, que a su vez representaban cuatro categorías: Validez, confiabilidad, impacto y ética. El instrumento puede leerse en su totalidad en el Anexo F. La siguiente tabla muestra la distribución de los reactivos entre las cuatro categorías:

Tabla 3.9*Clasificación de reactivos de encuesta por categorías*

Categoría	Número de reactivo	Total de reactivos
Aspectos éticos	34, 35, 36, 37, 38, 39, 42, 43, 44, 46, 47, 48, 49, 50, 51, 54	16
Confiabilidad	9, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 45, 69	24
Impacto	40, 41, 52, 53, 55, 56, 57, 58, 59, 70	10
Validez	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 21, 60, 61, 62, 63, 64, 65, 66, 67, 68	20
Suma		70

Se les pidió a los participantes que en el momento de contestar la encuesta seleccionaran la opción que representaba la frecuencia con que se llevaban a cabo la práctica mencionada en el reactivo. Las opciones eran:

- Muy frecuentemente (MF)
- Frecuentemente (F)
- Ocasionalmente (O)
- Casi nunca (CN)
- Nunca (N)

El instrumento se piloteo entre 17 profesores del departamento de exámenes del centro de lenguas de la universidad y dos profesores de lengua inglesa de un programa de licenciatura en la misma universidad. Entre los hallazgos del pilotaje, se encontró poca respuesta favorable ante la aplicación del instrumento vía digital. Solamente el 45% de los encuestados se dio a la tarea de contestar el cuestionario piloto. Se decidió que sería necesario hacer la aplicación de manera presencial y personal, con un instrumento en papel para asegurar la respuesta. Se tuvieron muchos comentarios sobre el orden de las preguntas, redacción y puntuación que permitieron mejoras en esos sentidos. Algunos de los

enunciados del instrumento parecieron repetitivos o confusos, aspecto que de igual manera se atendió.

Al hacer el análisis de los reactivos y por comentarios de los encuestados se redujo el número de preguntas de 128, en la primera versión del cuestionario, a 70 reactivos en su versión final. La extensión del instrumento se debió a la complejidad de los componentes teóricos que se estaba buscando explorar y las prácticas en las que se ven reflejados. La validez de un examen de lengua está relacionada con el constructo que es la competencia comunicativa, que a su vez tiene varios componentes y manifestaciones (Celce-Murcia, 2008) además de los distintos tipos de validez que son pertinentes en la práctica docente. De igual manera hay varias prácticas que incrementan o reducen el grado confiabilidad, de las que se consideró indispensable dar cuenta. Los aspectos éticos de la práctica resultaron ser numerosos y también era necesario considerarlos en el instrumento, así como las situaciones que podrían representar un impacto en el alumno. No obstante, el número de reactivos tuvo efectos en el análisis estadístico, específicamente en la proporción de ítems-participantes y en la prueba KMO, especialmente sensible también al tamaño de la muestra, como se podrá observar en una sección más adelante.

3.7.2 Instrumentación: Guion de entrevista semiestructurada.

En la segunda etapa se realizaron unas entrevistas semiestructuradas que nos permitieron conocer de manera más profunda y particular el pensamiento de los profesores. Se diseñó un guion para entrevista semiestructurada para indagar sobre las prácticas de evaluación y diseño de exámenes y profundizar en ellas sobre aspectos que el cuestionario no nos permitiría conocer, por su propia naturaleza, y consecuentemente, obtener un corpus que permitiría después realizar el análisis del discurso de los profesores y conocer sobre su responsabilidad profesional. El complemento entre la información obtenida con ayuda del cuestionario y la de las entrevistas nos permitiría tener una visión más completa sobre las prácticas de los diseñadores de exámenes y conocer la práctica docente con mayor amplitud.

Para diseñar el guion se establecieron una serie de objetivos, que a su vez fueron basados en aspectos que deseaban conocerse acerca de los participantes después de la encuesta, como prácticas detalladas de diseño de exámenes, experiencia en evaluación, percepciones

sobre las categorías teóricas como validez confiabilidad e impacto y aspectos éticos de la práctica evaluativa.

El guion se piloteó con entrevistas a dos profesoras de lengua inglesa de la universidad y una vez que se analizaron los resultados de este piloteo se llevaron a cabo algunos cambios con la intención de dar más fluidez y naturalidad a la entrevista y evitar respuestas repetitivas. Por ejemplo, se eliminaron las preguntas referentes a las relaciones académicas con los alumnos y aspectos del proceso enseñanza-aprendizaje, porque se consideró que se desviaban de los objetivos de esta investigación. Una vez que se obtuvo información acerca de las prácticas de evaluación y diseño de exámenes de lengua, se analizaron los datos de los cuales se lograron algunas inferencias y se redactaron nuevas preguntas que permitirían profundizar en las respuestas y obtener material para otros análisis. Por ejemplo, en el caso más evidente que fue la pregunta sobre el pilotaje de los instrumentos que elaboran los profesores, la mayoría manifestó no realizarlo. La entrevista nos permitió conocer las razones para esta omisión, entre otras prácticas, perspectivas y puntos de vista de los profesores. El guion resultante después del pilotaje y análisis, fue el siguiente:

Guion de entrevista

¿Dónde imparte clases a nivel licenciatura?

¿Cuántos años tiene(s) de experiencia como docente?

1. ¿Evalúa(s) a sus alumnos con exámenes? ¿Qué piensa(s) de evaluar a los alumnos de lenguas con un examen?
2. Podría(s) describir, paso a paso ¿cómo diseña(s) sus (tus) exámenes de lengua?
 - a. ¿Pilotea(s) sus exámenes?
 - b. ¿Evalúa(s) las cuatro habilidades?
 - c. ¿Evalúa(s) gramática y vocabulario?
3. ¿Qué características debe tener un examen?
4. ¿Alguna vez ha(s) dudado de un examen?
5. ¿Qué piensa(s) de sus exámenes?
6. ¿Ha(s) cambiado la forma en que diseña(s) exámenes de la primera vez que lo hizo (hiciste) a la fecha?
7. ¿Ha(s) recibido preparación técnica o específica en cuanto a diseño de exámenes?

8. ¿Qué tan importante es capacitarse en ese aspecto?
9. ¿Qué nos dicen sus (tus) exámenes de sus (tus) alumnos?
10. ¿Qué piensan sus (tus) alumnos de sus exámenes?
11. ¿Diría(s) que los exámenes ayudan o perjudican a los alumnos?
12. ¿Ha(s) reflexionado acerca del impacto que pudiese tener un examen elaborado por usted (ti) en la vida personal de un alumno?
13. ¿Qué nos dicen sus (tus) exámenes de ti como profesor de lengua?
14. ¿Qué puede (puedes) concluir con respecto al tema que hablamos hoy?
15. ¿Tiene(s) alguna pregunta sobre el tema o la investigación?

Las preguntas 2a, 2b y 2c se realizaron como preguntas cerradas ya que su objetivo era corroborar los resultados obtenidos en la fase cuantitativa, y eran seguidas de la pregunta “¿Por qué?” para conocer las razones de su práctica. Las preguntas 14 y 15 no estaban contempladas en el guion original, no obstante, durante la primera entrevista se pensó que eran preguntas lógicas y la información que se obtuvo fue muy enriquecedora, por lo que se decidió seguir haciendo las preguntas en las entrevistas subsecuentes.

3.8 Métodos de recopilación y análisis de información

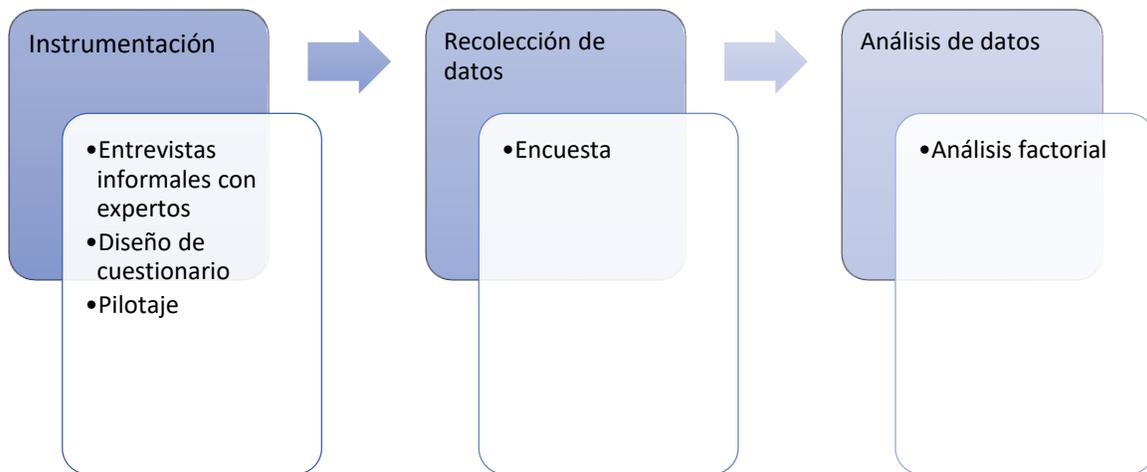
Una vez descritos los instrumentos de recopilación de información para las fases cuantitativa y cualitativa, se describirán los procesos de recolección de la información y su posterior procesamiento y análisis.

La fase cuantitativa se llevó a cabo en las siguientes etapas:

1. Instrumentación: Diseño de cuestionario
2. Recolección de datos: Encuesta
3. Análisis de datos:
 - a. Análisis factorial

Figura 3.1

Procedimiento de la fase cuantitativa de la investigación.



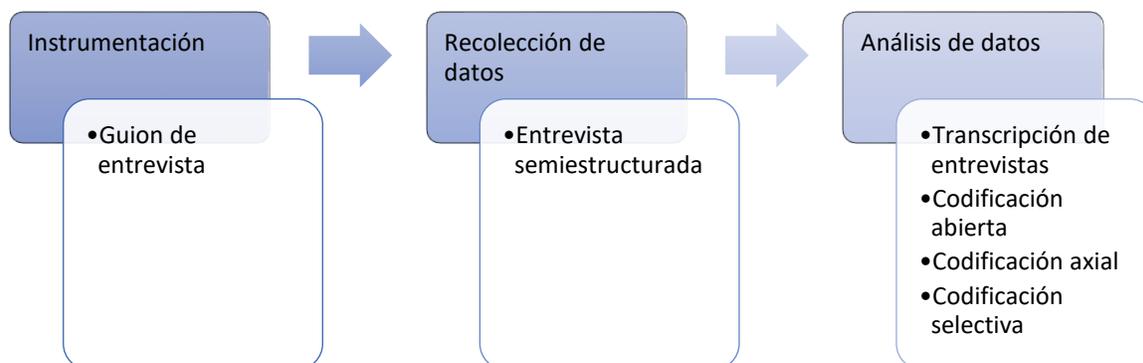
Nota: Elaboración propia

Por otra parte, la fase cualitativa consistió en las siguientes etapas:

1. Instrumentación: Diseño de guion de entrevista semiestructurada
2. Recolección de datos: Entrevista
3. Análisis de datos:
 - a. Transcripción de entrevistas
 - b. Codificación

Figura 3.2

Procedimiento de la fase cualitativa de la investigación.



NOTA: Elaboración propia

La fase cuantitativa se llevó a cabo en primera instancia. El proceso de instrumentación se describió a detalle en una sección previa. Los procesos de recolección y análisis de datos para ambas fases se detallan a continuación.

3.8.1 Recopilación de datos cuantitativos: Encuesta

La encuesta se realizó en un primer momento, a través de una comunicación vía correo electrónico con los coordinadores de lenguas de las facultades. Se les preguntó a estos por la manera en que se podría tener contacto con los profesores adscritos a sus organismos. Algunos se ofrecieron a aplicar el cuestionario ellos mismos, por lo que se les hizo llegar el instrumento de manera física. Esta forma de operar no representó una aportación importante en el número de participantes. Por el contrario, un par de coordinadores invitaron al investigador a asistir a una junta de trabajo con sus respectivos miembros de academia y donde se permitiría a este aplicar el instrumento directamente. En la Facultad de Lenguas, lugar de trabajo del investigador, se tuvo acceso a la información de los profesores y se les invitó a participar de manera directa y personal, además de la cercanía con él, lo que repercutió en un mayor número de participantes de este organismo. Se tuvo respuesta y participación de 13 de los 19 coordinadores, para lograr un total de 88 participantes. El instrumento se aplicó en versión en papel, dado el poco éxito que se tuvo con el piloto de manera electrónica.

3.8.2 Análisis de datos cuantitativos

Para analizar los datos recopilados a través de la encuesta, se decidió obtener un puntaje para cada reactivo. Este se obtuvo asignando un valor numérico a cada una de las opciones de respuesta, de la siguiente manera:

Tabla 3.10

Valores para opciones de respuesta de cuestionario

Opción	Valor asignado
Muy frecuentemente (MF)	5
Frecuentemente (F)	4
Ocasionalmente (O)	3
Casi nunca (CN)	2
Nunca (N)	1

La información obtenida en el procedimiento descrito en el inciso anterior se capturó en una hoja de cálculo de Microsoft Excel, para después ser procesada mediante un análisis estadístico en el software SPSS.

Se decidió llevar a cabo un análisis factorial exploratorio, con la finalidad de identificar los factores más trascendentes en la práctica evaluativa y que pudiese contribuir a la identificación y descripción de su estructura. La identificación de factores en esta etapa también permite la triangulación con los datos obtenidos en el análisis cualitativo.

3.8.2.1 Análisis factorial exploratorio

El análisis factorial es una técnica de análisis estadístico que permite analizar las correlaciones entre las variables, todas en conjunto, para detectar estructuras internas, definir grupos de variables o reducir el número de estas (Méndez Martínez y Rondón Sepúlveda, 2012). A estos grupos de variables se les conoce como factores.

Existen dos tipos de análisis factorial, el exploratorio y el confirmatorio. El análisis factorial exploratorio (AFE) se realiza cuando no se busca comprobar una hipótesis (Ferrando y Anguiano-Carrasco, 2010), sino que se busca explorar los factores para desarrollar teoría; lo contrario al análisis factorial comprobatorio (AFC), donde el investigador busca comprobar una hipótesis con respecto al número de factores, patrones de relaciones entre las variables y los factores o a las relaciones entre los factores.

Al revisar la literatura se encontraron investigaciones donde se utilizó el análisis factorial confirmatorio: Estudios sobre equidad en evaluaciones de lengua (Yan, Cheng y Ginther, 2018); sobre los constructos a medir en TOEFL-IBT (Sawaki, Stricker y Oranje, 2009); y sobre validez de una prueba de comprensión oral (Cai, 2013) utilizaron esta técnica.

El análisis factorial se ha utilizado para investigar sobre la literacidad en evaluación de lenguas. El AFC se utilizó para estudiar las concepciones de literacidad en evaluación de lenguas y necesidades de capacitación de los profesores (Vogt, Tsagari y Spanoudis, 2020); mientras que en un estudio de Kremmel y Harding (2019) el AFE fue utilizado para desarrollar un modelo empírico de literacidad en evaluación de lenguas. En este estudio se encuestaron 1086 participantes involucrados en los procesos de evaluación de lengua, como profesores, evaluadores, investigadores, administradores, alumnos y padres de familia, con la finalidad de conocer las necesidades de literacidad en evaluación de estos distintos grupos, desde la realidad de los participantes y no desde la teoría existente. En el caso de estudio que se presenta en este trabajo se busca generar una teoría con respecto a la práctica evaluativa, también desde una perspectiva práctica, por lo que se ha optado por el análisis factorial exploratorio.

Este tipo de análisis tiene ventajas que se han considerado relevantes para el enfoque que se ha adoptado en el estudio del caso, a pesar de las debilidades que se han documentado (Méndez Martínez y Rondón Sepúlveda, 2012; Lloret-Segura et al, 2014; Ledesma, Ferrando y Tosi, 2019; Kremmel y Harding, 2019). El tamaño de la muestra es reducido para llevar a cabo un análisis estadístico que permita realizar generalizaciones, sin embargo, es del interés del investigador entender las prácticas de los participantes desde su propio punto de vista, de describir y explorar más que comprobar la teoría existente. Se

considera que los factores resultantes del análisis permitirán esto y sentarán la base para una visión más completa del caso.

3.8.3 Recolección de datos: Entrevista

La entrevista se llevó a cabo en el lugar de preferencia del participante. Se ofreció un espacio en una oficina del investigador, la cual se encontraba en un lugar apartado y en silencio y permitía la grabación sin interrupciones. Once de los doce participantes decidieron acudir a dicho espacio, salvo una de ellas, Alondra, quien prefirió ser entrevistada en una oficina en su domicilio particular, donde se llevó a cabo sin ningún contratiempo. Todas las entrevistas se grabaron utilizando la aplicación Grabador de voz y se utilizó un dispositivo electrónico (iPhone) para ello.

3.8.4 Análisis de datos cualitativos

Las grabaciones de las entrevistas fueron descargadas de la aplicación y transcritas, por medio de la aplicación gratuita, en línea, OTranscribe. Una vez transcritas las entrevistas, se procedió a su análisis. Para ello se utilizó el software para análisis cualitativo Atlas TI, diseñado particularmente para el desarrollo de Teoría Fundamentada (San Martín, 2014). El uso de este software permitió que el proceso de análisis se realizara conforme a esta tradición metodológica que dio inicio con la codificación abierta (Figura 3.1, páginas adelante) y conforme se avanzó con este proceso se llevó a cabo el de codificación axial, donde se encontraron relaciones entre algunos de los códigos determinados en la codificación abierta.

Durante la realización de estos procesos, se llegó a la saturación teórica, que marcó la culminación del análisis primario o codificación abierta. La saturación teórica se dio durante el análisis de las entrevistas 11 y 12, donde no se encontraron más códigos nuevos, por lo que se determinó detener el análisis.

3.8.4.1 Uso de software para el análisis de la información cualitativa

San Martín (2014) propone el uso de software para el análisis cualitativo y en particular el uso del Atlas TI para desarrollar teoría fundamentada. San Martín comenta que el uso de software para el análisis cualitativo ha sido motivo de crítica por parte de investigadores,

quienes argumentan la necesidad de una labor más artesanal. El argumento de San Martín en favor del uso de Atlas TI es que el diseño de este software es particularmente adecuado para la teoría fundamentada, de tal manera que el método se lleva a cabo superando la práctica tradicional de análisis de datos cualitativos.

El Atlas TI permite al investigador realizar los procesos de la metodología de manera cíclica, simultánea y con flexibilidad. En este software es posible realizar la codificación abierta mediante el análisis de una variedad de tipos de documentos, como textos e imágenes. Como marca la metodología, una vez que se obtienen los códigos el Atlas TI permite, de manera muy práctica, crear redes y conexiones como parte de la codificación axial, lo cual también otorga la oportunidad de visualizarlas gráficamente. Todas las codificaciones, categorizaciones y conexiones quedan a la vista y contribuyen a la codificación selectiva.

Es necesario precisar que el software no aleja al investigador de los datos. La minuciosidad en el análisis y la perspectiva y visión del investigador quedan plasmadas en los reportes que pueden generarse. Sin ese trabajo cuidadoso y detallado por parte del investigador, los datos simplemente no podrían interpretarse y la teorización de la información no sería posible.

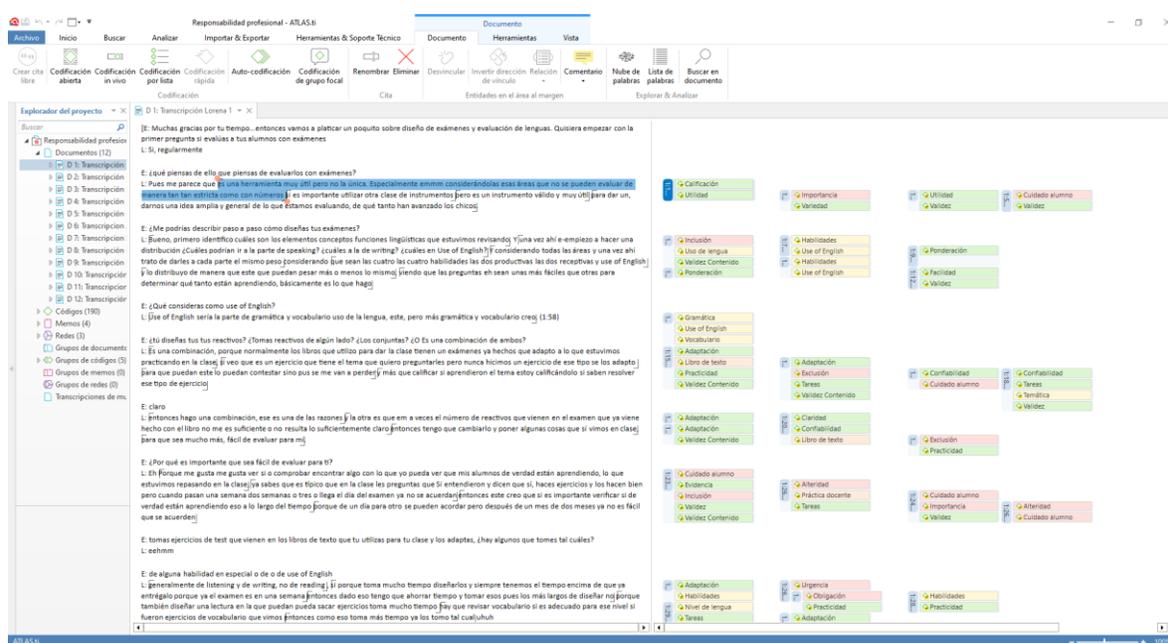
Dada la dimensión del estudio y la base metodológica de este software, la teoría fundamentada, se decidió utilizarlo como herramienta de análisis, al ser considerado por el investigador como una herramienta práctica, que permitiría el apego a la metodología elegida.

3.8.4.2 Codificación abierta

Se analizaron las transcripciones de las doce entrevistas a los participantes y cada extracto de ellas fue etiquetado con uno o más códigos. El proceso de codificación se llevó a cabo con ayuda de software Atlas TI. La siguiente figura (3.1) muestra cómo se realizó el proceso: Se seleccionó un extracto de la transcripción y se le asignaba un código. El software permite visualizar el extracto (o cita) y las etiquetas (códigos) asignados a esa cita.

Figura 3.3

Ejemplo de codificación abierta.



Nota: Imagen tomada de análisis de datos cualitativos en Atlas TI.

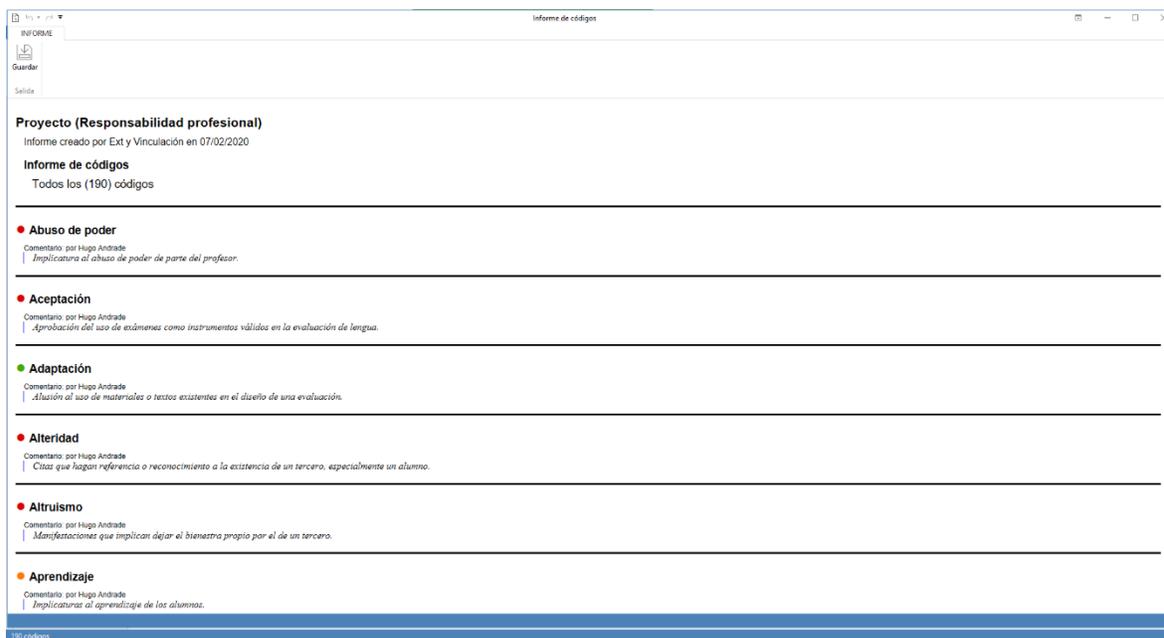
El Anexo S muestra un listado de todas las citas registradas, junto con los códigos asignados a cada una. Las citas se han organizado por números guía, o identificadores, compuestos por el número de participante y el número consecutivo para cada cita registrada en la entrevista de ese participante. Por ejemplo, la cita 1:14, corresponde a la participante 1 que es Lorena y el número 14 corresponde a la cita número 14 de la entrevista de ella. A lo largo del trabajo el lector encontrará extractos de las entrevistas, con el número de cita correspondiente.

Cada código nuevo recibía una denominación o etiqueta y una definición, de manera que cuando se volvía a utilizar un código se verificaba que la asignación de la etiqueta a ese nuevo extracto fuera congruente con la definición. Si no fuese así, se consideraba, ya sea, ampliar la definición o asignar un nuevo código al nuevo extracto.

A continuación, se muestra un ejemplo de algunas definiciones elaboradas para los códigos. La figura 3.2 es un extracto del reporte de los códigos y su definición, emitido por medio del Atlas TI y que se encuentra como el Anexo A.

Figura 3.4

Reporte de códigos (extracto).



Nota: Elaboración propia en Atlas TI.

En ocasiones, durante la codificación abierta, se realizaba una revisión de los códigos encontrados para ver si era posible unificar dos que fueran similares. Nuevamente se revisaba la definición y editaba en caso de ser necesario. Este proceso se realizó de manera constante durante el análisis.

Se realizó el proceso de codificación abierta para las doce transcripciones en su totalidad.

3.8.4.3 Codificación axial

Durante el proceso de codificación abierta dio inicio el proceso de codificación axial, donde a medida que emergían códigos del texto también eran notables algunas conexiones o relaciones entre ellos, las cuales se registraban en Atlas TI. Crear conexiones y encontrar relaciones entre los códigos impactó en el análisis en dos sentidos: Depuración de códigos y establecimiento de grupos de códigos.

Un ejemplo de depuración de códigos fue el siguiente. Durante la codificación abierta se encontraron extractos del texto que fueron codificados con la etiqueta Pilotaje, dado que se

referían a la acción de poner a prueba un examen antes de llevarlo a su aplicación, acción que en la práctica evaluativa repercute en un incremento del grado de confiabilidad de un examen. El código de confiabilidad se había asignado a algunos extractos de discurso, así que se decidió dejar de utilizar el código Pilotaje y fusionarlo con el código de mayor jerarquía, Confiabilidad, que a su vez engloba varias otras acciones con relación a esta cualidad de los exámenes. De esta manera, fue posible reducir el número de códigos de 224 a 190, en su mayoría códigos que se asemejaban o que se relacionaban tan estrechamente que podrían representarse con uno solo.

Una vez obtenidos los códigos y las dimensiones, se procesó un informe de cada uno de los participantes, con los códigos resultantes del análisis de la entrevista. Esto nos permitiría conocer acerca de la práctica evaluativa de cada uno de los participantes.

El proceso de codificación axial se complementó una vez que se concluyó con la codificación abierta de los segmentos de discurso de las entrevistas. Con ayuda del Atlas TI se obtuvo una tabla de co-ocurrencias (Anexo B), con un coeficiente-C de similitud entre los códigos. Este coeficiente muestra qué tan fuerte es la relación entre dos códigos, de manera parecida a un coeficiente de correlación, también muestra un índice entre 1.0 y 0.0, donde 1.0 muestra una relación más fuerte entre los códigos. La obtención de este índice permitió identificar las relaciones entre los códigos, paso fundamental en el proceso de codificación axial. A diferencia de un índice de correlación, como coeficiente de Pearson, para el coeficiente-C no hay valores que representen una relación fuerte entre dos códigos. Para fines de exploración de las relaciones entre códigos, se determinó identificar aquellas relaciones que mostraran un valor superior al índice de 0.06 y explorar las citas para determinar si la relación entre los códigos podría ser explicada.

Se procesó en el software Atlas TI una matriz que muestra las categorías de manera horizontal y vertical para identificar el número de veces que las categorías coincidieron con otras, en las citas. Es decir, la tabla matriz muestra el número de veces que la categoría Validez se encuentra en la misma cita que la categoría Confiabilidad, esta coincidencia significaría una relación entre ambas categorías. Se ha denominado a esta, tabla de co-ocurrencias, la cual permite visualizar las correlaciones más importantes entre todas las categorías, lo que podría no ser posible dada la cantidad de categorías.

De igual manera, se obtuvo una tabla matricial donde se muestran los códigos y su distribución entre los participantes (Anexo C). Esta tabla nos permite observar cuáles son los códigos con mayor presencia en el discurso de cada uno de los participantes. De manera similar, se presenta una tabla que muestra los códigos o categorías para cada dimensión (Anexo D). Esta tabla es necesaria para determinar la presencia de cada categoría en cada una de las dimensiones determinadas.

Se consideró necesario realizar las tablas mencionadas para determinar cuáles eran los códigos con mayor presencia o impacto en el discurso de los participantes.

El uso del software Atlas TI permite el registro de las relaciones de los códigos y fue también posible emitir un reporte de los 190 códigos encontrados, con los códigos relacionados. Este reporte puede verse en el Anexo E. La siguiente figura (3.3) es un extracto de dicho reporte:

Figura 3.5

Lista de vínculos código-código (Extracto).

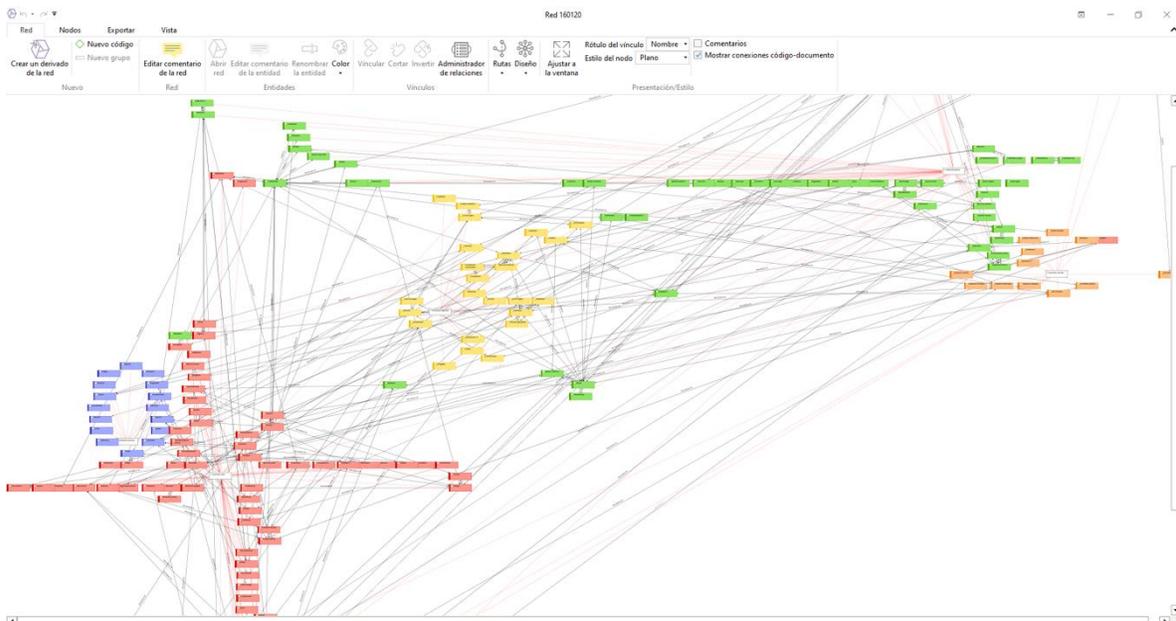


De igual manera, para visualizar gráficamente, se generó un diagrama que muestra las dimensiones, con sus respectivos códigos y las relaciones que se encontraron entre estos. Esto se realizó con la ayuda del software Atlas TI, que permite la presentación de las denominadas en el programa redes. La Figura 3.6, abajo, muestra la red correspondiente al análisis de las 12 entrevistas. Los colores corresponden a cada una de las dimensiones, estos colores fueron asignados por parte del investigador a los códigos, con base en la dimensión bajo los cuales se agruparon.

- Dimensión Técnica: Verde
- Dimensión Ética: Rojo
- Dimensión Lingüística; Amarillo
- Dimensión Docente: Naranja
- Dimensión Personal: Azul

Figura 3.6

Red de Dimensiones y códigos relacionados.



Nota: Imagen tomada de análisis en Atlas TI.

En esta figura (3.4) se pueden observar las conexiones entre los códigos y las categorías. La posición o ubicación de éstos en la figura provee una idea general de cómo se han construido dichas conexiones.

3.8.4.4 Codificación selectiva

Una vez que se han agrupado los códigos en categorías, estas a su vez conforman una categoría central, en un nivel de abstracción que permite explicar los datos encontrados. San Martín (2014) se refiere a este proceso no como independiente de la codificación abierta o axial sino como una extensión de ellos.

Para la conformación de este código central se consideraron los códigos encontrados en la codificación abierta, las dimensiones conformadas en la codificación axial y las relaciones entre los códigos y las dimensiones. El código central es la generación de la teoría que debe ayudar al entendimiento de la estructura de codificación realizada en el proceso de análisis y el papel del investigador es procesar los datos obtenidos del análisis, mediante un proceso de reflexión continua que permita lograr una explicación y conclusión satisfactoria al fenómeno de estudio.

En este caso, el proceso descrito ha llevado a la generación del código central que ha sido denominado *responsabilidad profesional*. En capítulos posteriores se explica con detalle cómo se llegó a esta conclusión del proceso, así como sus implicaciones y relaciones con la teoría existente.

3.9 Rigor metodológico

La investigación cualitativa es cuestionada, en ocasiones, debido a la falta de datos que puedan confirmar la validez o confiabilidad de la información presentada. Estos términos, generalmente relacionados con la tradición cuantitativa, deben ser reformulados. Flick (2011) recomienda documentar todas las decisiones tomadas por el investigador y hacer una clara distinción entre la información emitida por los participantes y las interpretaciones por parte del investigador. Flick también recomienda revisar las entrevistas para identificar algún tipo de comunicación entre los participantes y el investigador que resultara inadecuada para los fines de la entrevista.

En este trabajo se intenta dar cuenta, con el mayor detalle posible, de los criterios que se consideraron en la toma de decisiones. De igual manera, se buscó que toda la información quedara disponible para su confirmabilidad y contribuir a la credibilidad en el potencial lector. El criterio de transferibilidad se cumple con el establecimiento de teoría sustantiva que permita ser aplicada en otros contextos similares (San Martín, 2014).

A lo largo del proyecto de investigación y a la luz de los hallazgos se fueron tomando una serie de decisiones de corte metodológico y teóricas que contribuyeron a llegar al punto donde se encuentra en este momento. No obstante, hubo momentos en los que se tomaron rutas que desviaron el cumplimiento de los objetivos.

Desde el inicio del proyecto, la idea de estudiar la práctica evaluativa de los docentes desde un enfoque mixto era clara. En el momento de entrar en la fase cualitativa, el estudio se abordó con una idea preconcebida de lo que se iba a encontrar en el discurso de los participantes, con unos códigos predeterminados que se ajustaban a los postulados y teorías encontradas en la formulación del marco teórico. Este error metodológico pudo ser corregido gracias a las observaciones de las lectoras del trabajo de investigación en el reporte parcial.

En ese momento, se decidió buscar dentro de las tradiciones de la metodología cualitativa la que ayudara a contestar las preguntas de investigación y contribuyera al logro de los objetivos y que a la vez pudiera dar cuenta del procedimiento metodológico y del rigor aplicado en la obtención de los hallazgos. La teoría fundamentada fue la tradición que permitiría dicha rendición de cuentas. Esta decisión permitió que los datos fuesen, por sí mismos, los que determinaran los hallazgos y las respuestas a las preguntas planteadas por el investigador. Se analizaron las entrevistas con la premisa de dejar que los datos fluyeran y conformaran lo que sería la categoría central resultante.

CAPÍTULO IV

ANÁLISIS DE RESULTADOS

Introducción al capítulo

En este capítulo se presentan los resultados de las dos fases del proyecto, cuantitativa y cualitativa.

La investigación tuvo como objetivo general deconstruir la práctica evaluativa del profesor de lengua como diseñador de exámenes de lengua extranjera. A continuación, se presentan paso a paso los resultados obtenidos que han permitido el cumplimiento de este objetivo.

En la fase cuantitativa fue posible mostrar, de manera general y a través de la determinación de factores, la constitución de la práctica evaluativa de los participantes en el contexto descrito. En seguida, se dará cuenta de cómo, a través de la metodología de la teoría fundamentada, fue posible identificar la responsabilidad profesional y describir cómo se constituye en el discurso de los profesores, dando así cumplimiento al segundo elemento del objetivo general.

Los objetivos particulares coadyuvaron en la consecución del objetivo general. La encuesta dio la oportunidad de conocer las prácticas de los profesores, en cuanto a diseño de exámenes se refiere. Con base en el análisis factorial, fue posible conocer los factores que determinan las decisiones que llevan al profesor a dichas prácticas. Finalmente, a través del análisis del discurso se pudo averiguar cómo el profesor da cuenta de su práctica en diseño de exámenes.

El cómo se dio logro de estos objetivos se detalla a continuación, con la descripción del análisis de los datos cuantitativos y cualitativos.

ANÁLISIS DE DATOS CUANTITATIVOS

4.1 Fase cuantitativa: Encuesta

Como fase inicial de este estudio y con la finalidad de conocer las prácticas evaluativas de los profesores universitarios de lengua, se decidió desarrollar una encuesta que se aplicó a 88 sujetos. En seguida se muestran los resultados obtenidos de esta fase del estudio.

Con base en los puntajes asignados a las opciones de respuesta, se calcularon los porcentajes para el total de los reactivos y para cada una de las categorías como se muestra en la siguiente tabla:

Tabla 4.1

Porcentaje de respuestas por categoría inicial.

Categoría	MF	F	O	CN	N	NC
Aspectos éticos	53.1	26.4	9.8	5.9	4.3	0.4
Confiabilidad	32.8	31.6	17.1	10.4	7.7	0.4
Impacto	50.0	37.7	9.5	2.0	0.7	0.0
Validez	40.1	37.1	16.0	4.5	1.4	0.9
Total	39.3	31.2	14.1	8.1	6.9	0.5

Nota: Elaboración propia

La tabla nos muestra el porcentaje de personas que respondieron cada una de las opciones y divididos en categorías. Por ejemplo, el 53.1% eligió la opción MUY FRECUENTEMENTE (MF) como respuesta a las prácticas que reflejan los aspectos éticos del profesor. Esta opción de respuesta es la que se acerca más a la práctica deseable. La opción NUNCA, por el contrario, se aleja de lo que se esperaría de un diseñador de exámenes.

De la tabla podemos concluir que la mayoría de las respuestas se acercan más al extremo que representa las prácticas deseables del diseñador. Considerando el total de las respuestas, entre MF y F cubren el 70.3%, lo cual indica que, desde su propia percepción, los profesores y diseñadores de exámenes realizan su práctica en cuanto a diseño de exámenes y evaluación correcta o conforme a lo esperado de ellos. En contraste, solamente el 15% de las prácticas serían cuestionables, en cierto grado.

Al analizar el comportamiento de las respuestas por categoría se puede percibir que el aspecto ético es el que más importancia tiene entre los docentes y diseñadores de exámenes, con un 53.1% que contestaron MF, el mayor número para cualquier categoría, y 26.4 contestaron F, para una suma de estas dos respuestas de 79.5%, inclusive mayor al porcentaje mostrado en el total de las respuestas, de 70.3%. Esto indica que el profesor diseñador de exámenes se desempeña teniendo el aspecto ético como una prioridad en su práctica evaluativa. Aunado a lo anterior, se puede observar también una importante conciencia de parte del profesor del impacto que tienen la evaluación y los exámenes en los alumnos, mostrando entre las dos respuestas MF y F el porcentaje mayor, 87.7%, para cualquier categoría.

Por otro lado, la categoría que muestra mayor dispersión entre sus respuestas es la Confiabilidad. Esta categoría solo tuvo un 32.8% en que se contestó MF y un 31.6% para F. La suma de estas dos hace un total de 63.6%, el más bajo para cualquier categoría. Este dato nos indica que existe una mayor relajación en cuanto a cuidar aspectos de confiabilidad en las prácticas evaluativas. Es preciso aclarar que no se manifiesta una reducida confiabilidad, sino que es menor la atención de los docentes hacia este aspecto que hacia los demás.

En una apreciación preliminar, podría sugerirse que el docente tiene mayor consideración del aspecto ético de su práctica que de otros, como la confiabilidad. Esta situación podría parecer paradójica puesto que un instrumento de evaluación no puede ser válido si no es confiable, y si no es válido su uso para la toma de decisiones sobre la competencia lingüística es cuestionable.

No obstante, las afirmaciones en el párrafo anterior tienen matices. Los resultados no muestran que las prácticas evaluativas y diseño de exámenes no sean confiables, sino que es el aspecto que tiene menor influencia en las prácticas de los participantes de las categorías mostradas. Aun así, el 64.4% de las respuestas ubican las prácticas evaluativas de los docentes dentro de un parámetro deseable.

4.1.1 Análisis de datos

Una vez aplicada la encuesta y habiendo realizado el procesamiento inicial de los datos, se realizaron dos análisis: en primer lugar, estadística descriptiva para explicar el comportamiento de los datos cuantitativos obtenidos; y en segunda instancia un análisis factorial exploratorio, que nos permitiría observar el comportamiento de los datos manera conjunta, a través de la determinación de grupos de variables o factores.

4.1.1.1 Análisis de estadística descriptiva

Se llevó a cabo una estadística descriptiva en Microsoft Excel, en primer lugar, del total de los datos y después por categoría. La siguiente tabla muestra los resultados de este análisis:

Tabla 4.2

Estadística descriptiva para el cuestionario de prácticas evaluativas y diseño de exámenes

Categoría	Media	Mediana	Moda	Desviación Estándar	Varianza	Max	Min
Ética	4.187	4.165	4.727	0.483	0.233	4.76	3.16
Confiabilidad	3.662	3.737	3.534	0.710	0.504	4.61	2.23
Impacto	4.343	4.392	4.409	0.151	0.023	4.52	4.07
Validez	4.109	4.229	4.432	0.480	0.231	4.81	3.20
Total	4.007	4.125	4.727	0.594	0.353	4.81	2.23

Nota: Elaboración propia

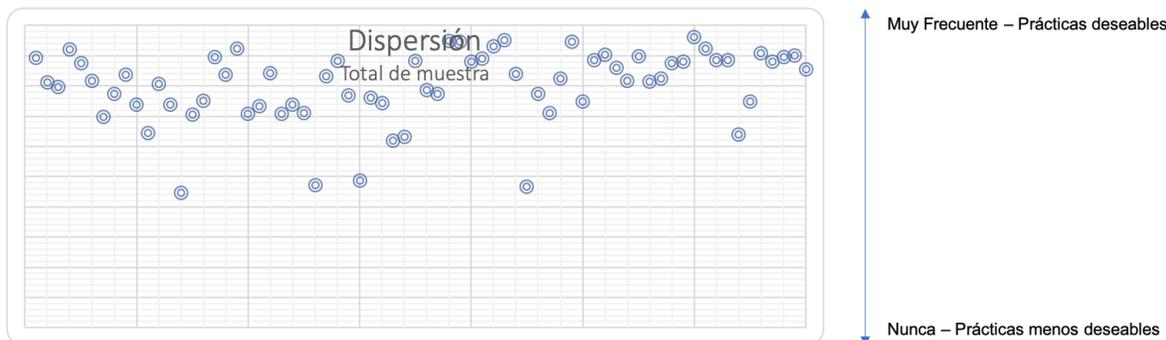
Se puede observar que la confiabilidad tiene la media más baja de las cuatro categorías con 3.662, mucho más abajo que la media general, que es 4.007. Sobre la confiabilidad debe resaltarse también que su desviación estándar, así como la varianza, son las más grandes entre las categorías, lo que indica una mayor dispersión de datos; tal como se muestra en la columna de los valores mínimos, siendo la confiabilidad la categoría que muestra el valor mínimo con 2.23. La baja significativa en la práctica relacionada con la confiabilidad de los instrumentos puede deberse a la pregunta sobre si realizaban un pilotaje de sus exámenes, a la que todos los participantes contestaron NUNCA.

El análisis de la estadística descriptiva confirma la importancia que le dan los docentes y diseñadores de exámenes a los aspectos éticos y de conciencia de impacto de sus instrumentos en la vida de los estudiantes. Ética e Impacto son las dos categorías con las más altas medias, con 4.187 y 4.343 respectivamente. Son notables las bajas desviación estándar y varianza que muestran los datos acerca de la categoría Impacto, con 0.151 y 0.023 respectivamente, esto significa que hay poca dispersión entre las respuestas; en otras palabras, la media en esta categoría es alta debido a que las respuestas se aglutinan principalmente entre las opciones MF y F.

En la siguiente gráfica se puede apreciar que, en general, los promedios de las respuestas se presentan en la parte superior de la gráfica, considerando que el número cinco fue asignado como valor a las opciones más deseables en la práctica evaluativa y de diseño de exámenes y en grado descendente a las prácticas menos deseables, que se acercan más a la parte inferior de la tabla. Cada punto de la tabla representa el puntaje promedio de cada reactivo del cuestionario. Por ejemplo, los cuatro puntos que se presentan más abajo en la tabla corresponden a los promedios de puntaje de los reactivos 14, 26, 30 y 45, coincidentemente todos pertenecen a la categoría de confiabilidad.

Gráfica 5.1

Dispersión de datos de cuestionario (Elaboración propia).



Nota: Elaboración propia

Lo que se entiende de lo anterior es que en la percepción que tienen los profesores de sus propias prácticas, la mayoría se ubican dentro de las prácticas deseables de un profesional

de evaluación de lengua y que a primera vista parece cumplir con los requerimientos de un profesional de la evaluación y el diseño de exámenes de lengua. El profesor, en su generalidad, manifiesta que en el momento de diseñar sus instrumentos y evaluar a sus estudiantes cumple con las prácticas deseables de su profesión, sus instrumentos son válidos, actúan con un alto grado de ética profesional, y una alta conciencia del impacto que tienen sus instrumentos en sus alumnos. La gráfica de dispersión nos muestra las distintas categorías de análisis pero con base en la estadística descriptiva se puede notar que los profesores reconocen que no cumplen con todos los aspectos necesarios para dar confiabilidad a sus instrumentos, al menos no al grado en que reconocen cumplir con el resto de las variables.

4.1.1.2 Análisis Factorial exploratorio

Después de la estadística descriptiva, se procedió a realizar el análisis factorial exploratorio, cuyo proceso de cálculo se describe a continuación.

Una vez asignado el valor numérico a cada respuesta se obtuvo el puntaje promedio para cada uno de los reactivos para su análisis gráfico y estadístico. Estos datos se cargaron en el software de análisis estadístico SPSS para obtener Alfa de Cronbach y realizar la prueba de interrelación de Keyser-Meyer-Olkin (KMO). Una vez realizadas estas pruebas se llevó a cabo el análisis factorial exploratorio.

Se calcularon los valores alfa de Cronbach para medir los niveles de consistencia interna de los reactivos. Como valor alfa, se obtuvo ,874, $N=59$, lo cual indica que los resultados de 59 reactivos correlacionan para generar una consistencia que da confiabilidad en el instrumento.

Logrado ese nivel de confiabilidad, se procedió a la prueba KMO cuyo resultado fue ,519, que apenas rebasa el límite de 0,5 como el mínimo requerido para realizar un análisis factorial exploratorio. La explicación posible para este resultado es el número de reactivos del instrumento (59 ítems después de reducción por alfa de Cronbach) que pudiesen ser demasiados para el reducido tamaño de la muestra (88 participantes). Dada la situación de recolección de datos, se consideró pertinente realizar el análisis. Para el estudio de este caso se contó con el mayor número posible de participantes que representan un 48.9% del

universo (N=182), con un 95% de confianza y margen de error de 7.45%, con lo cual se decidió trabajar. Con estas reservas, se procedió a realizar el análisis factorial exploratorio como se explica a continuación.

El análisis factorial exploratorio se realizó con máxima verosimilitud como método de extracción y Varimax como método de rotación. Para determinar el número de factores, se consideraron los seis que explican el 41.621% del total de la varianza, como puede verse en la siguiente tabla:

Tabla 4.3

Análisis factorial exploratorio: Varianza total explicada

Autovalores iniciales			
Factor	Total	% de varianza	% acumulado
1	9,629	16,320	16,320
2	3,717	6,301	22,621
3	3,489	5,914	28,535
4	2,804	4,753	33,288
5	2,504	4,243	37,531
6	2,413	4,090	41,621

Debido al número de factores con valores superiores a 1.0, no fue posible utilizar el método de determinación de factores de Kaiser. Este método sugiere utilizar aquellos factores cuyos valores *eigen* sean mayores a 1.0, en este caso fueron más de 17 valores, que son demasiados, aunado a la desestimación de dicho método debido a imprecisiones relacionadas con los tamaños de las muestras y con el número de factores obtenidos después de simulaciones (Navarro Loli y Merino Soto, 2010).

Los seis grupos de variables encontrados pueden observarse en la matriz de factor rotado (Anexo T), donde pueden observarse las columnas pertenecientes a cada factor obtenido y en las filas los ítems del cuestionario que resultaron agrupados en los factores.

De este modo, se conformaron los posibles factores con base en el análisis y que se muestran en las tablas a continuación, donde se muestran los ítems componentes:

El factor número 1 representa los valores y el impacto del examen. Los reactivos presentados son los que mostraron una mayor correlación entre sí y se refieren a la congruencia tripartita entre los valores y objetivos individuales, sociales y del sistema educativo y la conciencia del impacto de los exámenes y las consecuencias que los resultados de las evaluaciones tienen en el mismo sentido, así como la conciencia de responsabilidad sobre la validez del instrumento y las consecuencias que acarrea su carencia.

Tabla 4.4

Factor 1: Valores e impacto

Número de ítem	Ítem
55	Existe congruencia entre el propósito del examen y mis valores y objetivos.
56	Reflexiono en las consecuencias positivas o negativas para el alumno, la sociedad y el sistema educativo cuando utilizo un examen para algún propósito en particular.
57	Las interpretaciones sobre los resultados del examen son congruentes con los valores y objetivos de la sociedad y el sistema educativo.
58	Mis valores y objetivos son congruentes con los de la sociedad y del sistema educativo.
59	Me aseguro que las consecuencias que tenga el resultado de un examen no sean ocasionadas por un problema de validez del instrumento que diseñé.

El factor 2 corresponde a acciones relacionadas con responsabilidad durante diferentes etapas del proceso de evaluación.

Tabla 4.5*Factor 2: Responsabilidad en el proceso de evaluación*

Número	Ítem
25	Los exámenes se diseñan con base en normas y teoría de medición.
32	Los reactivos de un examen son revisados, por al menos una persona distinta, antes de su inclusión en la versión final a aplicar.
39	Estoy consciente de mi responsabilidad por los resultados del instrumento que he diseñado.
44	Hago una autocrítica de mi proceso de diseño de exámenes.
51	Estoy consciente de que al ejercer una evaluación a mis alumnos también tengo una responsabilidad hacia los otros miembros de mi academia.
52	Me aseguro que la retroalimentación del examen ayude al alumno a mejorar su habilidad lingüística.
53	Realizo evaluación continua con retroalimentación.
63	Complemento la evaluación que realizo con exámenes con otros tipos de evaluación como continua y auténtica, auto y co-evaluación.
49	Trato con respeto, sensibilidad e igualdad al alumno durante el proceso de evaluación.

Las prácticas descritas en los enunciados agrupados en este factor implican procesos de evaluación del aprendizaje de lengua realizados de manera responsable, que implican trabajar con una base teórica y técnica, en revisar los instrumentos elaborados en colaboración con otros profesores, realizar diferentes tipos de evaluaciones, retroalimentación efectiva y conciencia sobre la responsabilidad que implican sus acciones como evaluadores. Se decidió agrupar todos estos aspectos bajo el factor Responsabilidad en el proceso de evaluación.

El tercer factor corresponde a aspectos técnicos en el diseño de instrumentos de evaluación. Estos aspectos están relacionados particularmente con el concepto de confiabilidad. Las acciones descritas en los reactivos se relacionan con la búsqueda de confiabilidad al relacionar las tareas de un examen, uso de tablas de especificaciones y claves de respuesta, uniformidad entre exámenes aplicados a diferentes grupos de alumnos, así como preparación sobre aspectos técnicos de diseño de instrumentos y su aplicación.

Tabla 4.6

Factor 3: Aspectos técnicos

Número	Ítem
8	Las tareas a resolver en los exámenes están interrelacionadas.
9	El diseño del examen está basado en una tabla de especificaciones establecida.
12	Aplico conocimiento técnico de diseño de exámenes cuando elaboro un instrumento.
13	Me preparo en cuanto a procedimientos técnicos de diseño de exámenes.
15	Me aseguro que haya uniformidad entre los exámenes diseñados por mí y por otros profesores.
22	Los exámenes cuentan con una clave de respuestas detallada, que contempla todas las posibilidades de respuesta.

El cuarto factor agrupa reactivos que se enfocan en el propósito y los objetivos que tienen los exámenes, como claridad y definición en el propósito, tomar objetivos como base y usos que pueden darse al examen como que sirvan como ayuda al logro de los objetivos y reconocimiento del aprendizaje de los alumnos.

Tabla 4.7

Factor 4: Propósitos y objetivos del examen.

Número	Ítem
1	Las secciones de los exámenes que diseño ponen a prueba habilidades integradas.
40	El examen ayuda a reconocer los aspectos positivos del aprendizaje de los alumnos.
41	El examen ayuda a verificar el cumplimiento de objetivos del curso.
60	Diseño los exámenes basándome en los objetivos del curso o de la unidad.
61	Establezco claramente el propósito de un examen desde el inicio de su diseño.

El quinto factor agrupa cuatro variables que se agruparon en aquellas prácticas que contribuyen a incrementar la utilidad del examen, o *test usefulness* en términos de Bachman y Palmer (1996). Los reactivos agrupados se refieren a la validez, confiabilidad e interactividad de los exámenes.

Tabla 4.8

Factor 5: Utilidad del examen.

Número	Ítem
5	Las tareas a realizar en el examen son adecuadas a las características específicas de los alumnos, como edad, intereses, etc.
27	Los exámenes se preparan con suficiente tiempo antes de la fecha de aplicación.
62	En los exámenes busco evaluar conocimiento.
21	Los exámenes contribuyen a hacer notar claramente la diferencia entre los alumnos con habilidades menos desarrolladas de los que tienen habilidades más fuertes.

Finalmente, el factor 6, agrupa algunos aspectos sobre la preparación del examen, como piloteo y capacitación de los profesores como examinadores de producción oral.

Tabla 4.9

Factor 6: Preparación del examen.

Número	Ítem
26	Los exámenes se pilotean antes de ser aplicados.
14	En caso de llevarse a cabo un piloteo del examen, se modifican los reactivos con base en un análisis de los resultados.
30	Recibo capacitación como examinador oral.

Los seis factores obtenidos del análisis factorial exploratorio son:

Tabla 4.10*Análisis factorial exploratorio: Factores resultantes*

Factor	Descripción
1	Valores e impacto
2	Responsabilidad en el proceso de evaluación
3	Aspectos técnicos
4	Propósitos y objetivos del examen
5	Utilidad del examen
6	Preparación del examen.

Una vez realizados estos análisis, se determinó la necesidad de profundizar en las razones que llevan a realizar sus prácticas evaluativas y de triangular la información. Se considera que los factores obtenidos después del análisis se justifican teóricamente. Es decir, fue posible encontrar denominadores en común entre los ítems agrupados en el AFE. Con este análisis fue posible cumplir con dos de los objetivos planteados que fueron “Examinar las prácticas del profesor de lengua en diseño de exámenes de lengua extranjera” y “Determinar los factores de la práctica evaluativa en que se basa el profesor para la toma de decisiones”.

Los análisis presentados arriba nos otorgaron una base de inicio para la conformación de los perfiles de responsabilidad profesional y para entender el porqué de algunas de las manifestaciones resultantes del estudio en su fase cuantitativa. No obstante, para dar cuenta de la deconstrucción de la práctica evaluativa y cumplir en mayor grado con el objetivo general, se procedió al análisis de datos cualitativos.

ANÁLISIS DE DATOS CUALITATIVOS

En este punto del trabajo, se han presentado los datos estadísticos descriptivos y los resultantes del análisis factorial exploratorio, ambos de la fase cuantitativa del estudio. A continuación se presenta el análisis de los datos cualitativos obtenidos de las entrevistas a 12 participantes en el estudio, lo que permitió lograr el objetivo de “identificar de qué manera el discurso del profesor de lengua da cuenta de su práctica en diseño de exámenes”.

4.2 Fase cualitativa: Entrevistas

En este apartado del estudio se presentan una serie de tablas y figuras que servirán de base para comprender el proceso evaluativo descrito por los profesores en su discurso. A pesar de la naturaleza cualitativa de este análisis, se presentan datos cuantificables como números de incidencias o de co-ocurrencias entre los códigos que surgieron en el análisis. La justificación para este procedimiento radica en la cantidad de datos, de segmentos de análisis y de códigos que de otra manera no sería posible estudiar. La cuantificación permite llevar la atención a algunos aspectos que destacan entre el resto y guiar su estudio y análisis. En los siguientes incisos, se presentan las tres etapas de análisis con base en la metodología de la teoría fundamentada: las codificaciones abierta, axial y selectiva.

4.2.1 Codificación abierta

La codificación abierta consistió en analizar cada una de las doce entrevistas en su totalidad. Cada extracto de la entrevista se analizó y se describió mediante la asignación de uno o más códigos. Este procedimiento se describe a detalle en el inciso 4.3. En total se obtuvieron 190 códigos, los cuáles se presentan en el Anexo U, ordenados de acuerdo con el número de incidencias.

Todos los códigos fueron descritos y, con base en ello, se etiquetaron los segmentos de las entrevistas, asignando uno o más códigos. Las descripciones completas se encuentran en el Anexo F. En seguida, se muestran las de los principales códigos, aquellos con mayor número de incidencias encontrados en el análisis:

Confiabilidad: Menciones relacionadas con cualquier aspecto de confiabilidad del instrumento, del evaluador o del evaluado.

Alteridad: Citas que hagan referencia o reconocimiento a la existencia de un tercero, especialmente un alumno.

Exclusión: Citas que especifican la primera persona del singular o la tercera persona del singular o plural como sujeto.

Habilidades: Mención de alguna de las seis habilidades, en general o en particular.

Inclusión: Citas donde se utiliza la primera persona del plural para denotar pertenencia a un grupo o comunidad.

Cuidado del alumno: Citas que se refieren al bienestar del alumno.

Validez: Referencias a la conciencia sobre un objeto de medición.

Reflexión: Citas que hacen referencia a un pensamiento, consideración o meditación de parte del hablante.

Validez de contenido: Referencias a la consideración de algún contenido del curso como parte de una evaluación.

Tareas: Referencia a tareas o técnicas de evaluación.

Opinión: Expresiones que anuncian una opinión, específicamente con la expresión "yo creo".

Practicidad: Referencias a los recursos como tiempo o dinero necesarios en el diseño de una evaluación.

Gramática: Referencia nominal a la gramática.

Impacto: Citas que impliquen conciencia sobre la posibilidad de un impacto de la evaluación en el alumno.

Dificultad: Citas que expresan dificultad en la resolución o diseño de una evaluación.

Diálogo: Referencia a comunicación real o hipotética entre el hablante y el alumno.

Vocabulario: Referencia nominal al vocabulario.

Autoevaluación: Citas en las que el hablante hace un juicio sobre sí mismo o un aspecto de su práctica.

Juicio alumno: Juicio sobre la conducta o desempeño de un alumno.

Adaptación: Alusión al uso de materiales o textos existentes en el diseño de una evaluación.

Colaboración: Referencias a trabajo realizado junto con colegas.

Temática: Referencia al contenido de un curso o un examen como temas.

Nivel de lengua: Menciones de nivel de competencia comunicativa, nominalmente o con base en algún marco de referencia.

Preparación técnica: Referencias a cursos o preparación técnica específicamente en diseño de evaluaciones o exámenes de lengua.

Intención: Expresiones de intención, en especial con el verbo tratar.

Conocimiento: Referencias al conocimiento como objeto de la evaluación.

Obligación: Expresión de exigencia o deber.

Estas denominaciones y definiciones de los códigos permitieron evitar asignaciones de denominaciones repetidas, es decir, durante la codificación se identificaban y definían los códigos de manera que a medida que se avanzara en el análisis y se etiquetaran códigos con diferente denominación, pero que describieran lo mismo.

4.2.1.1 Relaciones entre códigos

Durante la asignación de códigos a los segmentos de las entrevistas, se notó que existían relaciones entre ciertos códigos. Estas relaciones se establecieron de dos maneras:

- Relación por incidencia
- Relación teórica

Se denominó relación por incidencia a aquella que se da entre códigos que se repiten constantemente en un segmento de análisis del discurso. Fue posible determinar el número de coincidencias con ayuda del software Atlas TI y con el reporte emitido llamado tabla de co-ocurrencias. Este puede encontrarse en el Anexo A. Identificar los códigos con más incidencias, y aquellos con los que se relacionan, es un indicio de cómo se conforma el discurso y permite al investigador seguir un hilo conductor al entendimiento de la práctica evaluativa, ya que, de esta manera, el análisis no únicamente descompone, sino que se busca encontrar interrelaciones que dan cuenta de la complejidad del proceso y la necesidad de la comprensión del conjunto y no de sus partes.

En la siguiente página se muestra una tabla de co-ocurrencias de los diez códigos con más incidencias en el análisis (Tabla 4.11) y como relacionan con otros códigos. Las columnas muestran los diez códigos con más incidencias, en orden alfabético, y las filas los mismos diez códigos y el cruce indica el número de incidencias de ambos códigos. Esta tabla matricial muestra, sin embargo, solo una parte de los códigos es un extracto del anexo B.

Tabla 4.11:

Tabla de co-ocurrencias de los diez códigos más con incidencias

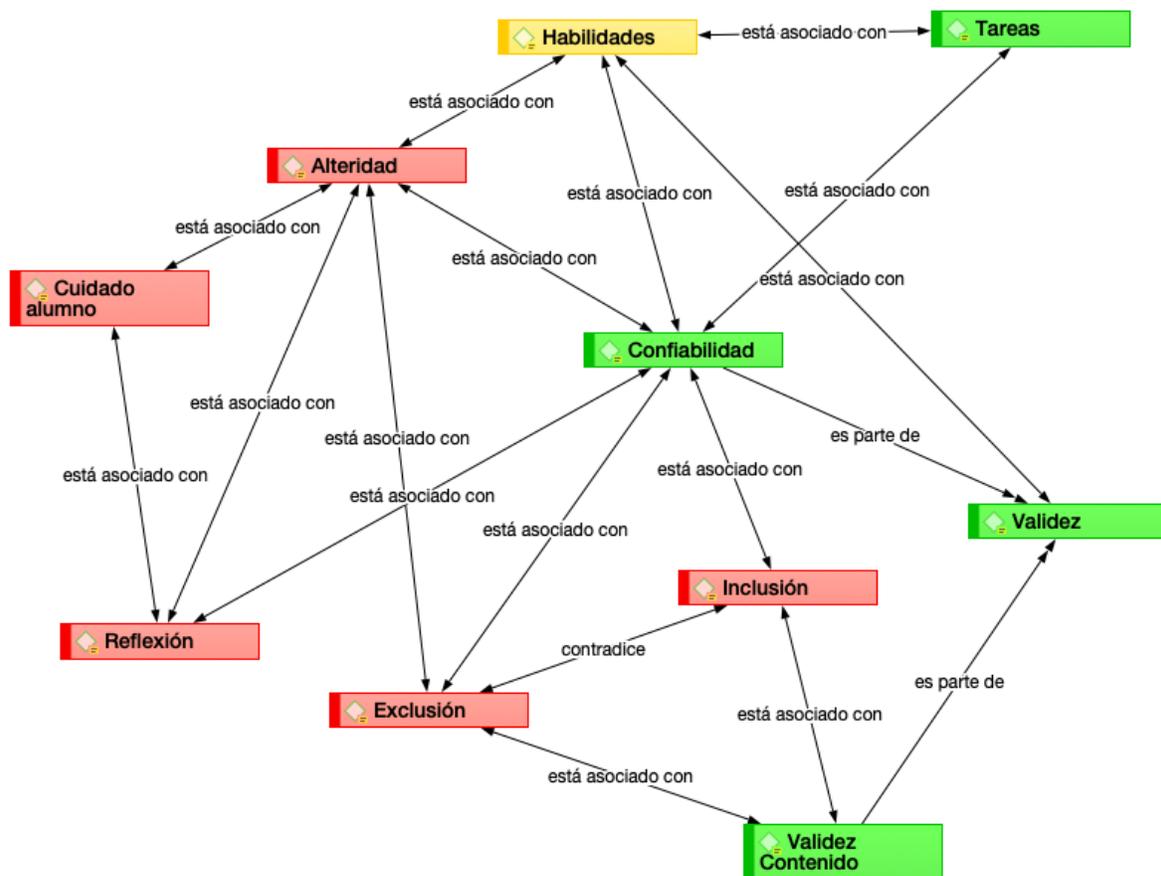
	Alteridad	Confiabilidad	Cuidado del alumno	Exclusión	Habilidades	Inclusión	Reflexión	Tareas	Validez	Validez de contenido
Alteridad	0	33	40	28	29	21	22	15	10	10
Confiabilidad	33	0	22	30	29	43	38	32	40	17
Cuidado del alumno	40	22	0	17	10	15	21	12	17	17
Exclusión	28	30	17	0	21	25	11	17	20	29
Habilidades	29	29	10	21	0	19	16	21	26	19
Inclusión	21	43	15	25	19	0	9	16	17	28
Reflexión	22	38	21	11	16	9	0	5	8	3
Tareas	15	32	12	17	21	16	5	0	13	11
Validez	10	40	17	20	26	17	8	13	0	8
Validez de contenido	10	17	17	29	19	28	3	11	8	0

Un ejemplo de relación por incidencias es el siguiente, el código con mayor número de incidencias, Confiabilidad, se relaciona principalmente por número de coincidencias con el código Inclusión, con el que coincide en 43 segmentos de análisis, en seguida con el código Validez, con 41 coincidencias, Reflexión con 38 coincidencias, Alteridad con 33 y Exclusión con 30. Esto da evidencia de una estrecha relación entre los códigos con mayor número de incidencias en el análisis.

La figura (4. 1) muestra los diez códigos con más incidencias y sus relaciones entre sí.

Figura 4.1:

Relación entre los diez códigos más con incidencias.



Nota: Elaboración propia en Atlas TI.

La relación teórica se refiere a las relaciones que se construyeron entre los códigos por sus conceptos, definiciones, descripciones o por similitudes en su naturaleza. Hay códigos que se relacionan porque representan aspectos teóricos y técnicos del diseño de exámenes,

como la validez, la confiabilidad y practicidad. Otros se relacionan al ser componentes de otros, como la validez y sus tipos, como validez de contenido y validez visual.

La codificación abierta, en resumen, ha permitido identificar 190 códigos que han sido definidos y clasificados con base en esta definición y sus relaciones con otros códigos.

Estas relaciones, por incidencia o por definición teórica, son las que darían lugar a los grupos de códigos que han sido denominados dimensiones y cuyo análisis y descripción se encuentra en la siguiente fase del análisis cualitativo denominada codificación axial y se presenta en el siguiente apartado de este trabajo.

4.2.2 Codificación axial

Como se mencionó en el inciso previo, en la codificación axial se crearon grupos de códigos. En el análisis de los códigos se detectó la emergencia de similitudes entre algunos códigos, aquellos que se enfocaban en el aspecto técnico del diseño de exámenes, o los que denotaban acciones que representaban una acción en beneficio de alguien, o los que se referían a algún aspecto lingüístico, como objeto de evaluación, entre otros casos. Este aglutinamiento de códigos resultó en cinco grupos, a los que se denominó dimensiones y que son las siguientes:

- Dimensión lingüística
- Dimensión técnica
- Dimensión ética
- Dimensión personal
- Dimensión docente

4.2.2.1 Grupos de códigos por dimensión

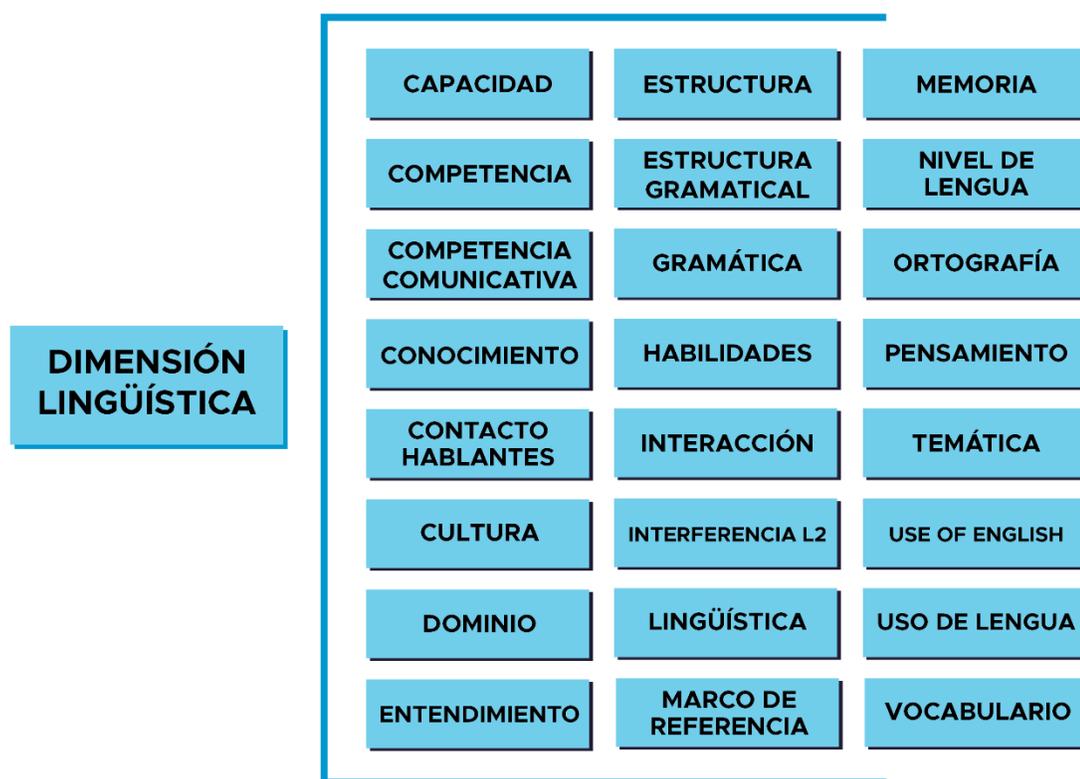
En seguida, pueden observarse los cinco grupos de códigos obtenidos del análisis, que se han denominado “dimensiones” y a cada una le acompañan los códigos agrupados en esa dimensión.

4.2.2.1.1 Dimensión lingüística

La dimensión lingüística agrupa todo aquello que emergió relacionado con las habilidades de la lengua así como aspectos cognitivos involucrados en todo proceso de aprendizaje. Se denominó así, en primer lugar, porque originalmente agrupaba los códigos que representaban un componente de la competencia lingüística o habilidades a ser medidos, como comprensión lectora, gramática o vocabulario. No obstante, al surgir en el análisis otros códigos, se decidió incorporar también aquellos que representaran algún aspecto a ser medido, aunque no fuese un componente de la competencia lingüística, como entendimiento, pensamiento o memoria.

Figura 4.2:

Dimensión lingüística



Nota: Elaboración propia.

4.2.2.1.2 Dimensión técnica

La dimensión técnica agrupa códigos con relación a algún aspecto técnico o del proceso del diseño de instrumentos de evaluación, como Validez, Confiabilidad, Practicidad o Autenticidad, denominados así por el metalenguaje de los principios y conceptos de evaluación. De igual manera, los aspectos relacionados, derivados o precedentes de éstos, como Tareas, Balance, Dificultad o Estandarización.

Figura 4.3:

Dimensión técnica

DIMENSIÓN TÉCNICA	ADAPTACIÓN	CONSTRUCTO	INTERACTIVIDAD	SIGNIFICATIVIDAD
	AUTENTICIDAD	CONTEXTUALIZACIÓN	INVERSIÓN TIEMPO	SISTEMATIZACIÓN
	BALANCE	COPIAR / PEGAR	LIMITACIÓN	TAREAS
	BANCO DE REACTIVOS	CREDIBILIDAD	MATERIAL AUTÉNTICO	TECNICISMO
	CALIBRACIÓN	DESARROLLO	METODOLOGÍA	TEORIZACIÓN
	CALIFICACIÓN	DIFICULTAD	ORGANIZACIÓN	TECNOLOGÍA
	CERTIFICACIÓN LENGUA	DISEÑO ORIGINAL	ORIGINALIDAD	UNIFORMIDAD
	CLARIDAD	ESPECIFICACIONES	PLANEACIÓN	UTILIDAD
	COLABORACIÓN	ESTANDARIZACIÓN	PONDERACIÓN	VALIDACIÓN
	COMPARACIÓN	EVIDENCIA	PRÁCTICA	VALIDEZ
	CONFIABILIDAD ALUMNO	EXPERIENCIA	PRACTICIDAD	VALIDEZ CONSTRUCTO
	CONFIABILIDAD	EXTENSIÓN EXAMEN	PREPARACIÓN TÉCNICA	VALIDEZ CONTENIDO
	CONOCIMIENTO TÉCNICO	FACE VALIDITY	PROBLEMA	VARIEDAD
	CONSTRUCTO	FACILIDAD	PROFESIONALIZACIÓN	SELECCION TEXTO
	CONFIABILIDAD ALUMNO	HOJA DE RESPUESTAS	REFERENCIA	INFLUENCIA
	CONFIABILIDAD	INCONGRUENCIA	REVISIÓN	CONOCIMIENTO TÉCNICO

4.2.2.1.3 Dimensión ética

La dimensión ética agrupa los códigos que implican un bienestar, propio para el participante, para algún individuo o grupos de ellos, o bien la falta de ello. Como ejemplo de los primeros, Obligación, Inclusión y Altruismo; de los segundos, códigos como Subjetividad o Engaño. Esta dimensión podría tener un par de subcategorías que se distinguen en el discurso y son la subdimensión interpersonal y la subdimensión valorativa, que serán explicadas más adelante.

Figura 4.4

Dimensión ética

DIMENSIÓN ÉTICA	ABUSO DE PODER	CONTRADICCIÓN	FAIRNESS	OBLIGACIÓN ALUMNO
	ACEPTACIÓN	CRÍTICA COLEGAS	FORMACIÓN ALUMNO	OPORTUNIDAD MEJORA
	ALTERIDAD	CUIDADO ALUMNO	IMPACTO	PRECAUCIÓN
	ALTRUISMO	CUMPLIMIENTO	IMPERSONALIDAD	PREJUICIO
	AUDACIA	DESCONFIANZA	IMPORTANCIA	PROFESIONALISMO
	AUTOCRÍTICA	DIÁLOGO	INCERTIDUMBRE	REFLEXIÓN
	AUTOEVALUACIÓN	DISCULPA	INCLUSIÓN	REFLEXIÓN ALUMNO
	AYUDA	DUDA	JUICIO ALUMNO	RESIGNACIÓN
	CALIDAD	EMPATÍA	JUSTICIA	RESPETO
	CAMBIO	ENGAÑO	JUSTIFICACIÓN	RESPONSABILIDAD
	COHERENCIA	ESENCIA	MEJORA	RETO ALUMNO
	COMODIDAD	ESFUERZO ALUMNO	MONETIZACIÓN	RETROALIMENTACIÓN ALUMNO
	COMPENSACIÓN	ESTEREOTIPOS	MOTIVACIÓN	SEGUIMIENTO ALUMNO
	COMPROMISO	ÉTICA PROFESIONAL	NECESIDAD ALUMNO	SENSIBILIDAD
	CONCIETIZACIÓN	EXCLUSIÓN	OBJETIVIDAD	SENTIMIENTO
	CONGRUENCIA	EXPECTATIVA	OBLIGACIÓN	SUBJETIVIDAD
	TOLERANCIA	TRABAJO	URGENCIA	UTOPIA
	VALORACIÓN	VALORACIÓN NEGATIVA	VALORACIÓN POSITIVA	

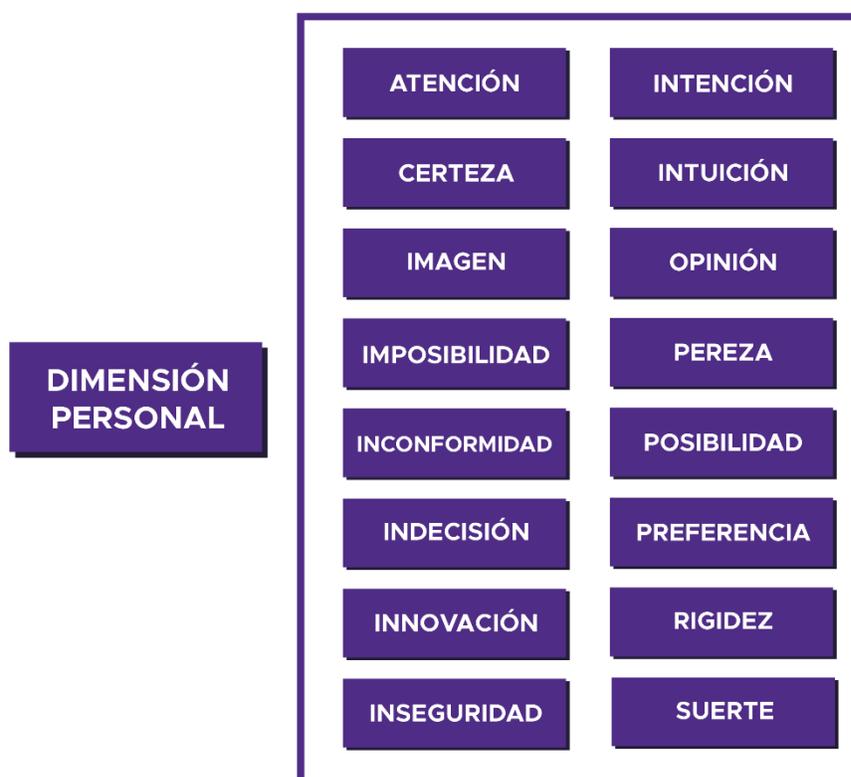
Nota: Elaboración propia

4.2.2.1.4 Dimensión personal

La dimensión personal agrupa códigos que muestran características, actitudes o posturas personales con respecto a los hechos presentados en el discurso. Estas no se relacionan con algún aspecto ético. Ejemplo de códigos dentro de esta dimensión son Opinión, Intención y Preferencia.

Figura 4.5

Dimensión personal



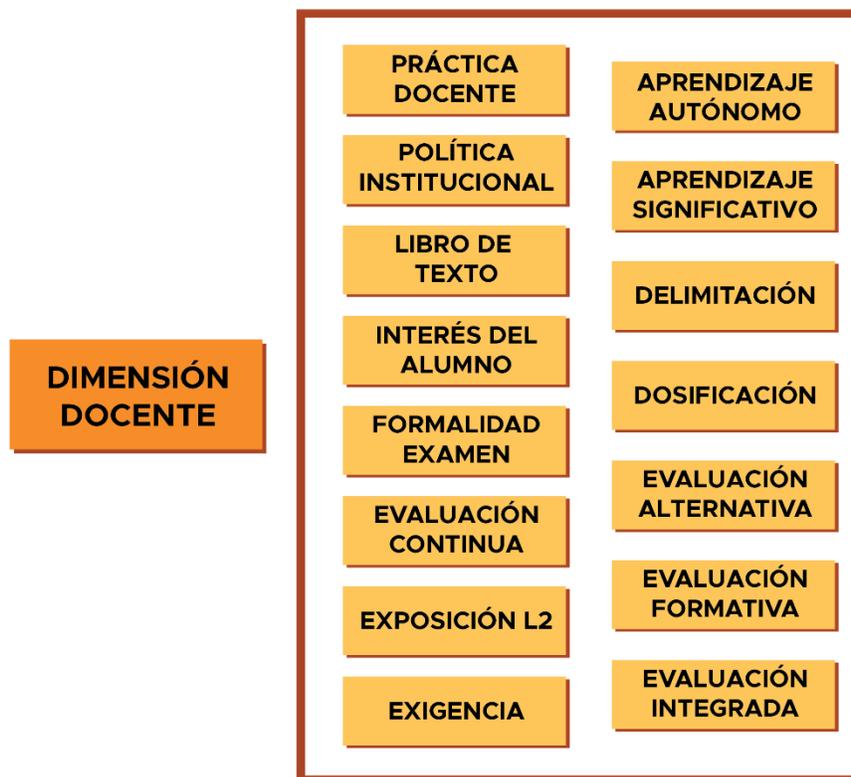
Nota: Elaboración propia.

4.2.2.1.5 Dimensión docente

La dimensión docente agrupa los códigos que hacen referencia a un aspecto de la práctica docente no relacionado directamente con el proceso de diseño de exámenes. Ejemplos de códigos dentro de esta dimensión son Libro de Texto, Práctica o Aprendizaje.

Figura 4.6

Dimensión docente



Nota: Elaboración propia.

Una vez determinadas las dimensiones y con la finalidad de tener una visión más amplia, se obtuvo una tabla matricial donde se muestran los códigos y su distribución entre los participantes (Tabla 4.3). Esta tabla nos permite observar cuáles son los códigos con mayor presencia en el discurso de cada uno de los participantes.

Tabla 4.12

Tabla de incidencias de códigos por participante (Elaboración propia, Atlas TI).

	Transcripción Lorena 1 Gr=251	Transcripción Erika 9.docx Gr=127	Transcripción Nora 10 Gr=205	Transcripción 5 Camila Gr=136	Transcripción Gina 11 Gr=98	Transcripción Alondra 6 Gr=233	Transcripción Simon 8 Gr=178	Transcripción Graciela 3 Gr=152	Transcripción Belinda 7 Gr=247	Transcripción Genaro 2.docx Gr=185	Transcripción Virginia 4 Gr=192	Totales	
1													
2	● Confiable Gr=356	43	32	21	31	9	88	27	17	38	20	30	356
3	● Alteridad Gr=271	41	13	50	14	11	8	37	36	17	17	27	271
4	● Exclusión Gr=210	23	8	22	12	9	16	35	15	17	17	36	210
5	● Habilidades Gr=205	19	9	41	13	5	25	21	9	12	30	21	205
6	● Cuidado alumno Gr=183	28	9	25	10	8	12	20	21	17	21	12	183
7	● Validez Gr=175	38	15	15	18	11	28	6	7	6	17	14	175
8	● Reflexión Gr=167	15	13	20	22	18	12	26	5	12	14	10	167
9	● Validez Contenido Gr=162	23	6	18	12	11	18	17	13	22	13	9	162
10	● Inclusión Gr=155	12	12	17	10	4	30	16	14	8	11	21	155
11	● Practicidad Gr=114	15	12	13	0	5	16	12	16	8	5	12	114
12	● Tareas Gr=109	12	7	12	9	0	9	20	8	0	10	22	109
13	● Gramática Gr=95	5	6	24	8	4	3	2	3	3	16	21	95
14	● Opinión Gr=93	28	15	13	8	0	0	0	13	0	0	16	93
15	● Vocabulario Gr=81	7	1	7	6	3	1	11	4	20	11	10	81
16	● Dificultad Gr=80	20	6	11	4	2	3	13	1	2	11	7	80
17	● Autoevaluación Gr=79	13	8	10	4	10	2	12	6	11	0	3	79

Nota: Elaboración propia, Atlas TI.

De manera similar, se presenta una tabla que muestra los códigos o categorías para cada dimensión (Tabla 4.4). Esta tabla es necesaria para determinar la presencia de cada categoría en cada una de las dimensiones determinadas. Se consideró necesario realizar las tablas mencionadas para determinar cuáles eran los códigos con mayor presencia o impacto en el discurso de los participantes.

Tabla 4.13

Tabla de incidencias de códigos por dimensión y participante (Elaboración propia, Atlas TI).

	Transcripción Lorena 1 Gr=251	Transcripción Erika 9.docx Gr=127	Transcripción Nora 10 Gr=205	Transcripción 5 Camila Gr=136	Transcripción Gina 11 Gr=98	Transcripción Alondra 6 Gr=233	Transcripción Simon 8 Gr=178	Transcripción Graciela 3 Gr=152	Transcripción Belinda 7 Gr=247	Transcripción Genaro 2.docx Gr=185	Transcripción Virginia 4 Gr=192	Totales
Dimensión docente Gr=161; GS=14	13	12	22	12	5	6	7	12	17	26	29	161
Dimensión ética Gr=1263; GS=72	151	74	159	88	73	112	141	116	119	117	113	1263
Dimensión lingüística Gr=522; GS=22	54	26	82	31	23	41	60	18	65	62	60	522
Dimensión personal Gr=219; GS=15	32	27	25	30	3	0	19	25	8	16	34	219
Dimensión técnica Gr=1269; GS=57	163	90	113	93	53	183	128	84	91	134	137	1269
Totales	413	229	401	254	157	342	355	255	300	355	373	3434

El uso del software Atlas TI permite el registro de las relaciones de los códigos y fue también posible emitir un reporte de los 190 códigos encontrados, con los códigos relacionados. Este reporte puede verse en el Anexo E. La siguiente figura (4.7) es un extracto de dicho reporte:

Figura 4.7

Lista de vínculos código-código (Extracto).

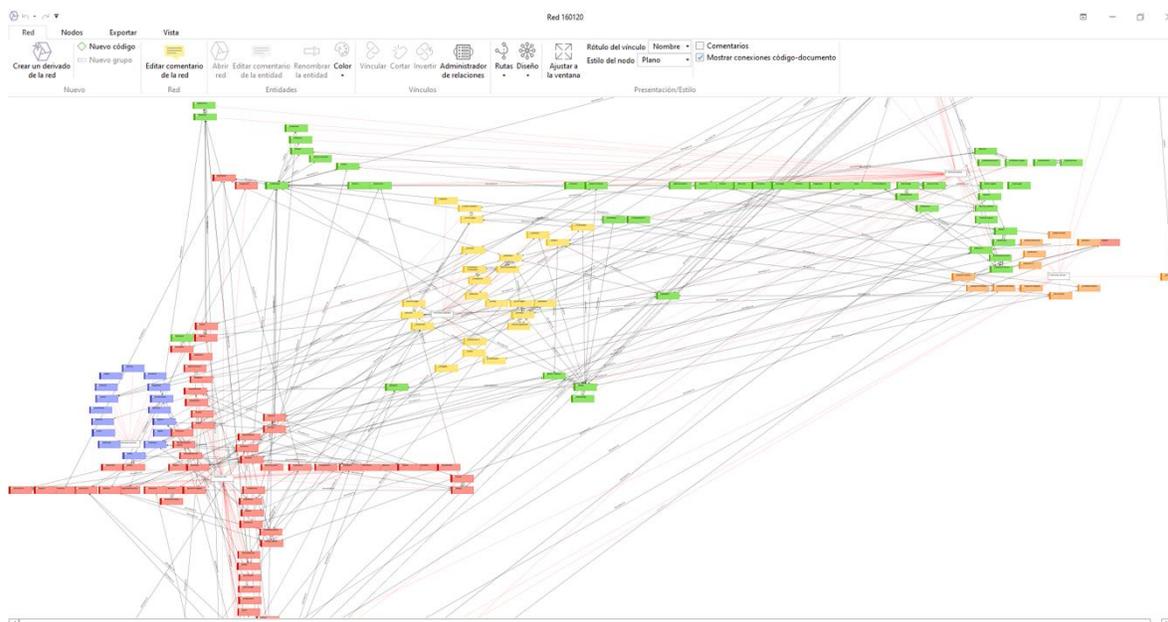


De igual manera, para visualizar gráficamente, se generó un diagrama que muestra las dimensiones, con sus respectivos códigos y las relaciones que se encontraron entre estos. Esto se realizó con la ayuda del software Atlas TI, que permite la presentación de las denominadas en el programa redes. La Figura 4.8, en la siguiente página, muestra la red correspondiente al análisis de las 12 entrevistas. Los colores corresponden a cada una de las dimensiones, estos colores fueron asignados por parte del investigador a los códigos, con base en la dimensión bajo los cuales se agruparon.

- Dimensión Técnica: Verde
- Dimensión Ética: Rojo
- Dimensión Lingüística; Amarillo
- Dimensión Docente: Naranja
- Dimensión Personal: Azul

Figura 4.8

Red de Dimensiones y códigos relacionados (Vista extendida)



Nota: Activar código QR para ampliar la vista de la figura 4.8 o activar hipervínculo:

https://drive.google.com/file/d/1gepFRsnnYt3fS6XxS_3hW9fBJgYbLZcz/view?usp=sharing

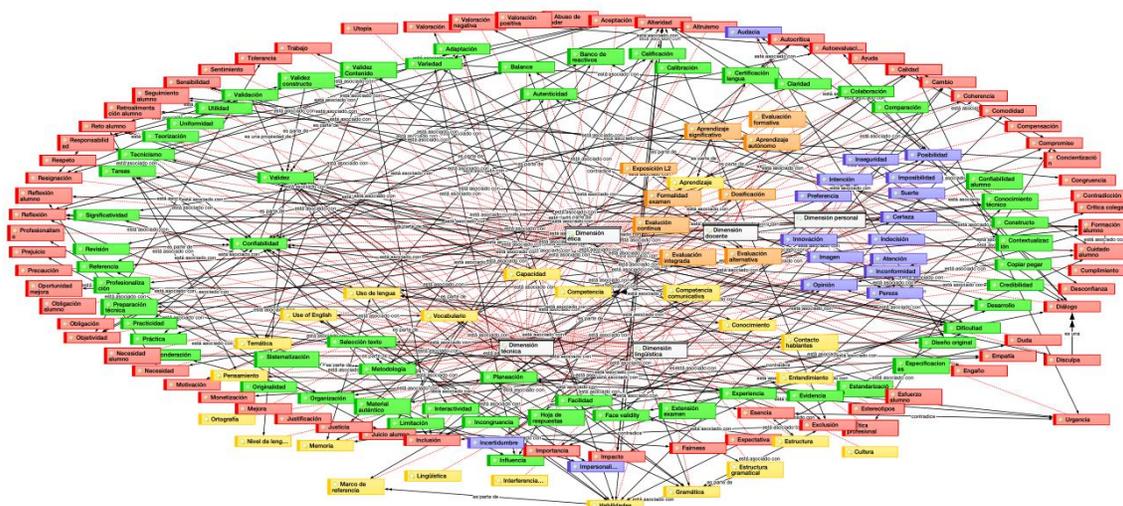
En esta figura (4.8) se pueden observar las conexiones entre los códigos y las categorías. La posición o ubicación de éstos en la figura provee una idea general de cómo se dan dichas conexiones. Puede observarse que la dimensión personal (en azul, en la figura 4.8) se encuentra cercana a la dimensión ética (rojo), alejada de otras dimensiones como la docente (naranja) y de la mayoría de los códigos de la dimensión técnica (verde). Puede percibirse también que la dimensión lingüística (amarillo) ocupa una posición central, lo que permite

observar sus conexiones con las dimensiones técnica y ética. La dimensión docente se encuentra en el otro extremo del diagrama con respecto a la dimensión personal, lo cual puede significar pocas conexiones entre ellas. Sin embargo, su cercanía a la dimensión técnica indica lo contrario, significativas conexiones entre sí. De igual manera se muestran varias conexiones entre la dimensión docente y la ética.

La siguiente figura (4.9) es otra presentación de las conexiones entre los códigos y las dimensiones:

Figura 4.9

Red de Dimensiones y códigos relacionados. Vista circular.



Activar código QR para ampliar la vista de la figura 4.9 o activar hipervínculo:

https://drive.google.com/file/d/1gepFRsnnYt3fS6XxS_3hW9fBJgYbLZcz/view?usp=sharing

En esta figura (4.9) generada en el software Atlas TI, es posible observar, en primera instancia, los códigos de la dimensión ética (rojo) y los de la dimensión lingüística (verde)

que por su mayor presencia en el discurso de los profesores se ubican en el exterior de la gráfica oval. Puede observarse una leve dispersión de los códigos de la dimensión lingüística (amarillo) y los códigos de las dimensiones personal (morado) y docente (naranja) aglutinados en el centro de la figura. Las líneas muestran las relaciones entre los códigos, entre los que destacan validez y confiabilidad, ambos de la dimensión técnica y que ocupan un lugar central en la figura debido a la cantidad de relaciones que se dan con otros códigos.

En síntesis, con base en la codificación abierta, la codificación axial ha dado oportunidad de clasificar los códigos en cinco dimensiones: ética, técnica, lingüística, personal y docente, de las cuales destacan las primeras tres. Se ha presentado la presencia de cada una de las dimensiones en el discurso de cada participante, así como ejemplos de las relaciones establecidas entre los códigos y que dieron origen a su clasificación en dichas dimensiones. Las relaciones entre ellos se presentan de manera gráfica, con el fin de ejemplificar lo estrecho de las conexiones y proveer una idea de la presencia de las dimensiones en el discurso de los participantes.

Así que una vez determinados los códigos en la codificación abierta y clasificados en la codificación axial, se procedió a la codificación selectiva, en la cual se hace una propuesta teórica que permite explicar la conformación del discurso de los participantes.

4.2.3 Codificación selectiva

Con base en los resultados obtenidos en la fase cuantitativa del estudio se tenía una idea de lo que se podría encontrar en la fase cualitativa, aunque con mucha mayor profundidad y en la voz de los mismos participantes. Los hallazgos en la etapa cualitativa que muestran un peso importante del aspecto ético en la práctica evaluativa, así como los aspectos técnicos principales, validez y confiabilidad, se encuentran con las cinco dimensiones identificadas en la fase cualitativa, técnica, ética, lingüística, personal y docente, para así conformar una dimensión central que se ha denominado **responsabilidad profesional**. Esta dimensión concentra todos los hallazgos obtenidos a través de las etapas del proyecto y la ubican como el hallazgo principal de este estudio. La responsabilidad profesional se comenta con más detalle en un apartado posterior.

La responsabilidad profesional es la capacidad de un profesional de la enseñanza de lengua para conjugar los factores técnicos, éticos, lingüísticos, docentes y personales para la toma de decisiones en su práctica evaluativa.

4.3 Análisis del discurso de los participantes

Una vez descritas las tres etapas de análisis de datos cualitativos, abierta, axial y selectiva, se presenta el análisis que se ha realizado del discurso de los participantes, las dimensiones y los códigos que se encontraron, así como evidencia de ellos en los textos.

4.3.1 La dimensión lingüística en el discurso de los profesores

En el estudio se encontró que la dimensión lingüística de los profesores se integra de 25 códigos, entre los que destacan seis: Habilidades, Gramática, Vocabulario, Temática, Nivel de Lengua y Conocimiento. La mayoría de los participantes en el estudio mostraron presencia de estos códigos en su discurso, con excepción de dos de ellos, cuyos discursos no presentan uno el código Temática y otro el código Conocimiento. En seguida se analizan estos y otros componentes menores de esta dimensión.

4.3.1.1 Habilidades, Gramática y Vocabulario

Como principales objetos de evaluación, los profesores entrevistados consideran las habilidades en general y de forma particular, por ejemplo, utilizan tanto “habilidades” o “habilidades receptivas”, como “comprensión lectora” o “*reading*”. De igual manera, se refieren a la gramática en general, “gramática”; a estructuras gramaticales, o a alguna en particular como “voz pasiva” o “*passive voice*”; o tiempos verbales, como “*passé composé*”. En cuanto al vocabulario, los profesores se refieren de esa misma manera, generalizando “vocabulario”. En seguida se muestra evidencia de esto.

En el siguiente extracto Lorena se refiere al proceso de diseño de sus exámenes y menciona las habilidades en general y su clasificación:

... considerando todas las áreas y una vez ahí trato de darles a cada parte el mismo peso considerando que sean las cuatro las cuatro habilidades las dos productivas las dos receptivas...
(Lorena: Entrevista, 1:9-10)

Erika comenta lo siguiente ante la pregunta de que si evalúa las cuatro habilidades.

Sí, sí viene *listening, reading, writing* y bueno *grammar. Speaking...* (Erika: Entrevista, 2:64)

Lorena hace referencia a las habilidades agrupadas (productivas y receptivas), mientras que Erika lo hace de manera particular, en inglés. Entre las habilidades Erika incluye la gramática (*grammar*), aunque matiza con la interjección “bueno”, lo que podría significar dos cosas, primero, que entiende que la gramática no es una habilidad y que se incluye, a pesar de ello, o, segundo, la obviedad de incluir la gramática en el examen.

Entre las habilidades, manifiestan evaluar gramática y vocabulario, códigos que entre sí se relacionan estrechamente en el discurso de los profesores. Todas las incidencias correspondientes a estos dos códigos se dieron de manera explícita.

Nora, al describir cómo evalúa a sus alumnos, se refiere a las habilidades, la gramática y vocabulario.

... y el examen escrito que es comprensión oral comprensión escrita, vocabulario, gramática y un poco de producción escrita... (Nora: Entrevista, 3:11)

Adicionalmente a la gramática y el vocabulario, se encontró, en menor medida, que los profesores se refieren, constantemente, como elementos de la dimensión lingüística, tanto a estructuras gramaticales de manera general, como a algunas en particular. Una muestra de cómo se utilizan las estructuras gramaticales para ejemplificar las prácticas de diseño de exámenes se encuentra en el siguiente extracto, en el que Genaro, al describir su procedimiento de diseño, menciona:

... es más fácil para mí encontrar ejercicios sobre sobre presente perfecto o sobre preposiciones... (Genaro, Entrevista, 10:54)

Genaro utiliza el presente perfecto y las preposiciones como ejemplo de cómo identifica estructuras o categorías gramaticales para sus exámenes de lengua.

Virginia pone énfasis en la evaluación de la gramática, pero presenta una excepción al resto de los entrevistados. La profesora manifiesta preferencia por evaluar las excepciones a las reglas gramaticales. Esto lo expresa de la siguiente manera:

A mí me gusta evaluar, por ejemplo, mucho las irregularidades. Tú sabes que toda regla gramatical tiene la regla ¿no? y sus irregularidades las excepciones. Yo evaluó mucho las excepciones, que se aprendan más la excepción a la regla. (Virginia, Entrevista, 11:33-34)

La presencia de estos tres códigos, Gramática, Vocabulario y Estructura gramatical, es evidencia de arraigo de la evaluación por punto diferenciado (*discrete point*) y del peso que tienen estos conceptos en la práctica docente y evaluativa de los profesores participantes en el estudio. Si bien, el concepto con más presencia es el de habilidades, la suma de estos otros tres conceptos equipara el de las habilidades.

4.3.1.2 Temática

Otro código que destaca es la temática. Este código se utilizó para señalar toda referencia al contenido de un curso de lengua que luego se considera como contenido en un examen. Los profesores se refieren a “temas estudiados en el curso” o a “temas” que contenga el examen, lo que resulta en una estrecha relación con la validez de contenido, elemento de la dimensión técnica.

Virginia, en su entrevista, menciona precisamente la “temática” al describir cómo diseña los exámenes:

... lo primero que hago es escuchar audios para ver cual tiene que ver con las temáticas que se abordaron en clase... (Virginia, Entrevista, 11:39)

Nadia, al opinar sobre la disyuntiva entre evaluar gramática y vocabulario de manera diferenciada o en habilidades integradas explica:

... trato de que a veces las frases que ellos tienen que completar o que tienen que responder las preguntas pues tengan que ver con el tema que vimos de la unidad para que este más contextualizado... (Nadia, Entrevista, 12:28)

Nadia utiliza la palabra “tema” para referirse al contenido del libro de texto, que es utilizado en una unidad para presentar los elementos lingüísticos en un contexto. Este “tema” también es utilizado en el examen que ella diseña para contextualizar los elementos a ser evaluados en el examen.

Si bien, la temática no es en sí un objeto de evaluación, sí es utilizada como base para ubicar al alumno en un contexto y así evaluar su competencia comunicativa o alguno de sus componentes. Esta contextualización se relaciona con la autenticidad, cualidad de los exámenes que a su vez se relaciona estrechamente con la validez de constructo.

4.3.1.3 Nivel de lengua

Por “nivel de lengua”, entiéndase cualquier referencia al nivel de competencia comunicativa. En algunos casos las referencias al nivel de lengua fueron nominales o con base en un marco de referencia, específicamente el MCRE. Al referirse al “nivel de lengua”, o al “nivel”, los profesores podrían referirse a cosas diferentes, por ejemplo, Lorena tiene como base el nivel para elegir el tipo de tareas a evaluar, así que de esta manera evalúa si la tarea es apropiada:

... hay que revisar vocabulario si es adecuado para ese nivel si fueron ejercicios de vocabulario que vimos... (Lorena: Entrevista, 1:29)

El nivel de lengua debe ser congruente con el examen, como menciona Nora al hablar de las características que debe tener un examen:

... yo creo que las características de un examen debe (*sic*) de ser del nivel de lengua, debe ser congruente con el nivel de lengua... (Nora: Entrevista, 3:41)

El nivel de lengua es un momento en el aprendizaje de los alumnos y con base en las declaraciones de los profesores, el examen es un instrumento que sirve para que los alumnos demuestren ese nivel o momento de aprendizaje de lengua en que se encuentran. El nivel de lengua sirve para que los profesores desarrollen instrumentos coherentes entre los contenidos de los cursos y los exámenes. Los contenidos generalmente indicados en los libros de texto, suelen ser coherentes con un marco de referencia, generalmente el MCRE, como se menciona en renglones previos. Cuando los profesores mencionan el nivel de lengua implican dichos contenidos y funciones como objeto de evaluación.

4.3.1.4 Conocimiento

El código Conocimiento implica referencias al conocimiento como un objeto de evaluación. En el análisis se consideró que el uso del verbo “saber” se relaciona con el sustantivo “conocimiento” que da nombre al código.

Graciela, al describir cómo diseña exámenes, expresa:

... entonces tomo lo que estuve viendo en clase, lo que considero que es necesario que los chavos sepan... (Graciela: Entrevista, 8:14-15).

Belinda utiliza la palabra “conocimiento” para referirse a lo que miden sus exámenes:

...el conocimiento que tiene de la lengua y cómo pueden usarla... (Belinda: Entrevista, 9:163).

Los profesores manifiestan evaluar el “conocimiento” o el “saber” de los alumnos con respecto a la lengua de estudio (Genaro, 10:88; Nadia, 12:16). De acuerdo con la literatura de evaluación y diseño de exámenes de lengua, y como se comentó previamente, el objeto de evaluación debe ser la competencia comunicativa y sus manifestaciones, es decir, las habilidades. La competencia comunicativa implica el uso del conocimiento en una función comunicativa, así que es necesario indagar si la manifestación sobre evaluar “conocimiento” o “saber” indica una visión parcial del objeto de evaluación en los exámenes de lengua y, como consecuencia, una necesidad de profundizar en el análisis de la competencia comunicativa por parte de los profesores. Es posible, también, que los profesores utilicen la palabra conocimiento como sinónimo de la competencia y no como un elemento de ella.

La dimensión lingüística es el “Qué” de la práctica evaluativa, y en el discurso de los profesores muestra su carácter esencial en la responsabilidad profesional. No se concibe una práctica profesional responsable sin una clara conciencia del objeto o constructo a evaluar, que para algunos será la competencia comunicativa, habilidades, contenidos del curso, nivel de lengua o conocimiento. No obstante, cada uno de los objetos mencionados representa un constructo o una serie de constructos complejos y que se interrelacionan entre sí, por lo que la definición de uno de ellos como objeto único de evaluación representa para los profesores

Independientemente de la denominación utilizada por los profesores, el conocimiento técnico de la evaluación de lenguas podría coadyuvar a que los profesores definan con más precisión y proveer de un grado más alto de validez a sus instrumentos, elemento esencial de los aspectos técnicos, que precisamente se explican a continuación en la dimensión técnica.

4.3.2 La dimensión técnica en el discurso de los profesores.

La dimensión técnica en el discurso de los profesores se compone de 60 códigos, siendo los más notables Confiabilidad, Validez, Validez de Contenido, Tareas, Practicidad, Dificultad, Adaptación, Preparación técnica, Tecnicismo, Colaboración y Calificación.

En este inciso se presenta la evidencia de la presencia de estos códigos en el discurso de los profesores. En primer lugar, se analizará el código Confiabilidad, por su complejidad, por el gran número de acciones y de componentes. Junto con este código se muestra el análisis de los códigos de menor presencia relacionados con este y los componentes que no se

consideraron como códigos en sí, sino que conforman el propio concepto de confiabilidad. En seguida, se analiza el código Validez y los códigos relacionados, como validez de contenido, y finalmente el resto de los códigos más importantes que conforman la dimensión técnica.

4.3.2.1 Confiabilidad

Como se ha visto previamente en este trabajo, la confiabilidad es la consistencia de resultados en un examen si se aplicara en dos momentos diferentes bajo condiciones similares. Este aspecto técnico del diseño de exámenes es el más presente en el discurso de los profesores. La confiabilidad a su vez requiere una serie de acciones y decisiones, las cuales se presentaron en este documento en su segmento teórico.

La confiabilidad es un concepto que está presente en los profesores, sin embargo, varios lo mencionan sin claridad acerca de su significado, como es el caso de Erika y de Genaro, como puede verse en los extractos que se presentan a continuación:

En una respuesta dada durante su entrevista Erika menciona el término confiabilidad, a lo que el entrevistador pregunta qué es. La respuesta de Erika:

¿Confiabilidad? ya no me acuerdo (Erika: Entrevista, 2:13)

A petición del entrevistador amplía su respuesta:

... Pues bueno a manera muy coloquial o muy vaga, yo creo que es este pues con un examen (*sic*) que te sentiste bien, que no tuviste como problema en aplicarlo, que no le encontraste detalles de redacción o sobre los temas o no te costó trabajo ¿no? (Erika: Entrevista, 2:14-15)

La explicación de Erika es imprecisa, aunque se acerca al concepto de confiabilidad.

Confiabilidad y tareas o técnicas de evaluación

Se encontró en el análisis del discurso de los profesores que la mayoría de ellos utiliza la palabra “ejercicio” para referirse a las tareas o técnicas de evaluación en un examen. Por ejemplo, en el siguiente extracto, la respuesta de Lorena a la pregunta si toma reactivos previamente diseñados de alguna otra fuente, contesta:

... si veo que es un ejercicio que tiene el tema que quiero preguntarles, pero nunca hicimos un ejercicio de ese tipo se los adapto para que puedan este lo puedan contestar sino pus se me van a perder y más que calificar si aprendieron el tema estoy calificándolo si saben resolver ese tipo de ejercicio... (Lorena: Entrevista, 1:16-18).

Además del uso de la palabra “ejercicio”, en esta cita se puede observar que Lorena puntualiza que la tarea debe ser conocida por los alumnos o de lo contrario estaría calificando su habilidad para resolver la tarea en lugar de lo que se debe calificar.

Camila utiliza la palabra *task* (tarea) al responder sobre cómo evalúa la producción escrita:

También la incluyo, les pongo las instrucciones del task, que ellos tienen que hacer, ellos lo realizan y yo parte pues tengo la rúbrica para evaluarlo... (Camila: Entrevista, 4:26).

Con excepción de Camila y de Gina y Belinda, quienes no hacen referencia a alguna tarea o técnica de evaluación, todos los participantes utilizan la palabra “ejercicio”. Esto implica poca familiarización con la terminología técnica.

Las referencias a algún tipo de tarea, como “juego de roles”, “composiciones” o “complementación” (Alondra, 6:3) también fueron categorizadas con el código Tareas.

Las entrevistas permitieron conocer cómo los profesores eligen las tareas a incluir en el examen. Como el caso de Nora, quien al describir su proceso de diseño de exámenes hace un comentario sobre el tipo de tareas incluidas en los libros de texto:

... la verdad es que muchos de los libros de texto, en los libros de profesor viene el disco con material extra y con exámenes personalmente no me gusta utilizar los ejercicios de gramática o de vocabulario de ahí porque son muy parecidos a los del libro, creo que no representan un reto para el alumno entonces yo trato muchas veces de incluir como otros tipos de ejercicios un poquito más complejos... (Nora: Entrevista, 3:17-18).

Nora decidió no utilizar tareas con las que estaban familiarizados los estudiantes, por considerar que no representaban un reto e intentó con un nuevo tipo de tarea en el examen, como se describe en los siguientes extractos:

... hace tiempo evalué a un grupo con un audio que describa, lo tome de un *podcast* pero la verdad es que me fue muy mal porque los chicos, pues sí fue material auténtico, entonces los chicos me decían estuvo súper interesante... (Nora: Entrevista, 3:17-18).

... a los chicos les gustó fue interesante pero no se concentraron en contestar las preguntas del examen... (Nora: Entrevista, 3:21).

... era un *fill in the gaps* y un *true and false* entonces se confundieron mucho, pero yo veía que estaban atentos, entonces no, les fue muy mal... (Nora: Entrevista, 3:22).

Nora hace una reflexión sobre el tipo de tarea y especialmente sobre el hecho que los alumnos no estaban familiarizados con ella. La tarea no funcionó y los resultados no fueron los que ella esperaba.

Lo anterior contrasta con lo que declara Alondra sobre cómo elige las tareas para su examen:

... elijo una lectura acorde al nivel, elijo tipos de ejercicios que se hayan practicado en clase, vamos con él, que los alumnos estén familiarizados ¿no? (Alondra: Entrevista, 6:74).

Alondra precisa que los alumnos deben conocer las tareas a realizar en el examen, que deben ser practicadas en clase para que se familiaricen con ellas.

Confiabilidad: Adaptación

En el diseño de exámenes de lengua, una práctica recurrente es adaptar tareas existentes y utilizarlas como parte del examen. Este código se presentó en todos los participantes en el estudio. Es posible observar evidencia de esta práctica en el discurso de los profesores en los siguientes extractos:

Camila se refiere a lo anterior cuando explica el procedimiento que sigue para diseñar sus exámenes:

... tomo de los libros, más bien material de los libros que nos proporcionan para el material de maestros, ciertos reactivos que me pudieran ayudar a conformar cada área del examen... (Camila: Entrevista, 4:21).

Gina expresa que obtiene las tareas de comprensión lectora de ejercicios de libros diseñados previamente.

Algunos ejercicios, bueno sobre todo las lecturas, esos sí los tomo de los libros... (Gina: Entrevista, 5:10).

No obstante, estas tareas pueden ser modificadas según determine la profesora:

En cuanto a las preguntas, si hay algunas que me convencen del ejercicio como tal las dejo, si hay otras cosas que me interesan que se pueden como aprovechar más, bueno, entonces cambio el ejercicio... (Gina: Entrevista, 5:12).

Se puede apreciar en la evidencia mostrada que los profesores utilizan y/o adaptan tareas de libros de texto para ser utilizadas en sus propios instrumentos.

Confiabilidad: Colaboración

Una manera de lograr confiabilidad en los instrumentos de evaluación es el trabajo colaborativo entre profesores. Cuando los instrumentos se diseñan con más de un punto de vista, o incluso algo mucho más sencillo como una revisión, se reduce la posibilidad de errores. En su discurso, los profesores se han expresado sobre este punto de la confiabilidad; Erika (Erika: Entrevista, 2:57) menciona que los exámenes se revisan entre los profesores de su centro de trabajo; Alondra (Alondra; Entrevista, 6:110) explica cómo sus instrumentos mejoraron gracias a las opiniones de otras profesoras; Lorena (Lorena: Entrevista, 1:111), por otro lado, expresa la dificultad que representa el trabajo colaborativo, principalmente por las condiciones de trabajo y aspectos como el tiempo; y en otra desventaja del trabajo colaborativo que surgió durante las entrevistas, Alondra (Alondra: Entrevista, 6:154) menciona los desacuerdos con sus colegas.

El discurso de los profesores muestra evidencia del trabajo colaborativo, que a su vez provee ventajas y desventajas, aunque mientras las diferencias parecen ser conceptuales, se podría concluir que en términos de confiabilidad los aspectos positivos del trabajo colaborativo representan una mejora en la práctica evaluativa.

Confiabilidad: Calificación

El código Calificación describe las citas en las que se hace referencia al proceso de calificación (*scoring*) de los instrumentos de evaluación. Sobre este tema hay posturas, prácticas, creencias disímiles como se muestra a continuación.

Para Simón los exámenes implican más que una calificación, cuando habla sobre cómo estos ayudan a los alumnos:

... no nada más se trata o al menos bueno como profesor yo lo veo de esta manera no solo se trata de la calificación, aunque pues para los alumnos es solamente la calificación, pero yo creo que a los alumnos les ayuda muchísimo porque tienen algo tangible que les dice cómo están o cómo salieron... (Simón: Entrevista, 7:144-145).

En contraste, Simón expresaría más adelante que los exámenes podrían tener un impacto negativo en los alumnos, justamente en cuanto a la calificación se refiere.

... yo creo que la única manera en que los perjudica es en calificación, en calificación me refiero pues bueno hay que asentar una calificación y ellos tienen que recibir su calificación y eso les da derecho quizá a becas... (Simón: Entrevista, 7:149-150).

En las dos citas anteriores, Simón parece tener una postura con respecto a la calificación, sugiere que es secundaria; en contraste, después menciona las implicaciones que tiene, al impactar en decisiones importantes que pueden afectar a los alumnos.

Graciela (Graciela: Entrevista, 8:106-107) expresa la urgencia a que se ven sometidos los profesores para obtener calificaciones, lo que tiene repercusiones en el diseño del examen y menciona que se le pide entregar calificaciones, el “subir calificaciones” se refiere a la plataforma en donde los profesores capturan las calificaciones en una fecha límite establecida. Graciela califica como “un lujo” un diseño de examen que no sea práctico para calificar.

Al hablar sobre la evaluación de la producción oral, Virginia (Virginia: Entrevista, 11:82-83) menciona que el cansancio podría ser un factor en la obtención de los resultados y afectar la objetividad.

Las calificaciones de los alumnos ayudan al docente a elaborar juicios que le ayudan a tomar decisiones, como comentan en este sentido Virginia (Virginia: Entrevista, 11:152) quien describe como al momento de calificar elabora juicios sobre los alumnos y describe las decisiones que toma con base en esas calificaciones y Nadia (Nadia: Entrevista, 12:15), quien además de expresar su gusto por los exámenes, de igual manera describe las decisiones tomadas con base en las calificaciones.

Calificar para los profesores representa una urgencia, pero a la vez representa un indicador sobre el desempeño del alumno y del mismo profesor en el diseño del instrumento y de la objetividad con que se evalúa a los alumnos.

4.3.2.2 Validez

La validez en términos generales es la cualidad de un examen de medir lo que se propone medir. Este concepto ha resultado central, debido a su relación con otros códigos de su misma dimensión, como con otras dimensiones, Validez se relaciona con la dimensión lingüística y con la dimensión ética, como puede observarse en la siguiente figura (4.5), en donde puede

observarse gráficamente la centralidad del concepto en todo el estudio. La gráfica muestra la validez que se ubica en el centro de la red, esto debido a las descritas relaciones con otros códigos y dimensiones.

En seguida se muestra el análisis del discurso y los hallazgos, en cuanto a este código, componentes de esta dimensión lingüística.

La validez aparece en el discurso de dos maneras, cuando se refieren a ella nominalmente y cuando se hace referencia al propósito de medir una característica; como Lorena (Lorena: Entrevista, 1:2-5) al hablar de los exámenes de lengua como instrumento de evaluación; Erika (Erika: Entrevista, 2:5-6) al hablar de las desventajas de los exámenes como instrumentos de evaluación; o Simón (Simón: Entrevista, 7:8-9), quien al contestar qué piensa de sus propios exámenes, expresa que “tiene que tener (*sic*) validez”.

La validez es un aspecto muy importante en la literatura de diseño de exámenes. En la práctica, algunos profesores tienen claridad sobre qué es, sin embargo, para otros no es así.

Erika mencionó el término validez, al preguntársele qué significaba contestó:

¿Validez? Yo entiendo que es que mida lo que tiene que medir ¿no? (Erika: Entrevista, 2:9).

Más adelante agregaría:

... para mí eso es validez que se haya utilizado otras veces y que se vea y se compruebe que los reactivos o bueno las preguntas están bien diseñadas. (Erika: Entrevista, 2:12).

En la cita 2:9 Erika concuerda con lo que se define en la literatura la validez. No obstante, en la segunda cita (2:12) parece confundir validez con confiabilidad.

Como se mencionó renglones arriba, las manifestaciones de conciencia de medir un aspecto del individuo en particular, fue considerada dentro del código Validez, el cual generalmente está acompañado de otro código que representa lo que se está midiendo o pretende medir, es decir, los aspectos sujetos de ser medidos con ayuda de los exámenes de lengua. Lorena, por ejemplo, expresa la intención de medir aprendizaje del alumno. En este extracto con el cuál concluye su explicación sobre cómo diseña sus exámenes de lengua, declara:

... viendo que las preguntas, eh, sean unas más fáciles que otras para determinar qué tanto están aprendiendo... (Lorena: Entrevista, 1:12).

También se consideraron como Validez las expresiones en que los profesores manifestaban coherencia entre lo que ellos buscaban con el examen y lo que se lograba, como muestra el siguiente ejemplo extraído de la entrevista a Camila:

... ahora ya me doy cuenta qué quiero que hagan los alumnos, que nada más identifiquen que analicen este reactivo o qué quiero que hagan con la información que les estoy dando... (Camila: Entrevista, 4:62-63).

En estos extractos pueden notarse una variedad de maneras en que los profesores se refieren o implican el concepto de validez. Adicionalmente a este concepto, visto desde un enfoque más general, también se refirieron o implicaron tipos de validez, como se explica en seguida:

Validez y validez de contenido

Algunos códigos se han considerado dentro de la dimensión técnica ya que contribuyen directamente a la validez. Por ejemplo, la validez de contenido, código que ha sido descrito como aquellas referencias a la consideración de algún contenido del curso como parte de una evaluación.

Al preguntarles a los profesores cómo diseñan sus exámenes, varios de los entrevistados (Lorena: Entrevista, 1:6; Nora: Entrevista, 3:14; Camila: Entrevista, 4:7; Gina: Entrevista, 5:4; Alondra: Entrevista, 6:74; Graciela: Entrevista, 8:12; Virginia: Entrevista, 11:39) expresaron considerar, en primer lugar, el contenido de los cursos.

Para Simón y para Gina, la validez de contenido es una característica de un “buen examen”. Así como lo expresan en los siguientes extractos:

Ok bueno un buen examen para mí sería un examen que está evaluando los temas que se hayan visto... (Simón: Entrevista, 7:80).

... si los preparé para algo, bueno pues se va a ver en el examen, tiene que ser muy coherente (sic) con lo que vimos... (Gina: Entrevista, 5:26).

La validez de contenido es relacionada con otros códigos, como justicia. Así lo expresa Nadia (Nadia: Entrevista, 12:110), quien menciona justicia y al preguntársele qué quiere decir que un examen sea justo, responde “Que sea justo significa que evalúes lo que les enseñaste”. Para Nadia, es justo evaluar lo que se les enseñó a los estudiantes durante el curso. Incluso, ve falta de lógica en la práctica contraria.

Virginia (Virginia: Entrevista, 11:71) también menciona el impacto que tiene en los alumnos la práctica de evaluar aquello que no se estudió en clase o la manera en que se hizo algo durante el curso. Relata una ocasión en que el examen que ella realizó tenía como tareas unas que no se habían practicado durante el curso.

Para Genaro, incluir en el examen algún aspecto que no se haya visto en el examen invalida el examen. Al hablar de la validez de los exámenes, explica cómo durante el curso omitió revisar ciertos ejercicios y el efecto en la evaluación:

... o en el caso de gramática cuando comenté que estábamos en un *workbook* por ejemplo y por ejemplo se pasó revisar la unidad y viene una pregunta en el examen sobre gramática entonces eso ya no lo hace válido porque ese conocimiento no lo tienen el alumno o más bien no se dio en clase... (Genaro: Entrevista, 10:87-88).

Graciela expresa haber equivocado su práctica evaluativa y haber perjudicado a los alumnos al no practicar un ejercicio que después sería una tarea que responder en el examen:

... una vez me pasó que no vimos el formato de ese ejercicio y les fue muy mal, entonces dije pues creo que ahí metí la pata yo porque no tuvimos práctica de ese ejercicio en particular entonces por eso creo que les fue tan mal... (Graciela: Entrevista, 8:57-59).

La validez de contenido también es justificación para utilizar exámenes prediseñados, como aquellos contenidos en los libros de texto. Así lo manifiesta Graciela al responder a la pregunta de por qué utilizaba dichos exámenes.

Ay se me hace muy tedioso (diseñar exámenes), pues es que como que siento que es lo que viene diseñado por lo que vimos o sea exactamente el vocabulario la gramática... (Graciela: Entrevista, 8:94).

Cabe mencionar que la validez de contenido es uno de los códigos que fueron encontrados en todas las entrevistas realizadas, lo cual es una confirmación de la importancia de este aspecto de la evaluación. Para los profesores la validez de contenido representa justicia, congruencia y coherencia en su quehacer, lo que es muestra de la confluencia de la validez y la ética. También son dignas de resaltar la riqueza y variedad de las formas en que los profesores se refieren al contenido de los cursos.

Validez y Autenticidad

El código Autenticidad se utilizó para describir las referencias al diseño de las tareas de un examen o una evaluación en las que el uso de la lengua sea lo más cercano a su uso real. La teoría sobre diseño de exámenes vincula la autenticidad de los exámenes de lengua directamente a la validez (Bachman y Palmer, 1996; Brown y Abeywickrama, 2010). Cualquier intento del profesor por buscar que el alumno haga un uso de la lengua, lo más cercano posible a la realidad, en un examen de lengua siempre tendrá un impacto directo en elevar el grado de validez.

En el estudio, los profesores se manifestaron acerca del uso de la lengua en exámenes. Algunos profesores, como Nora, mencionaron actividades comunicativas (Nora: Entrevista, 3:10) y que es necesario contextualizar las tareas y ella lo hace de manera complementaria a los exámenes (Nora: Entrevista, 3:12-13),

Alondra imparte clases a estudiantes de ingeniería y de medicina, en la siguiente cita puede verse cómo los contextos y necesidades específicas de cada facultad generan una reflexión en la profesora, quien explica que las necesidades de lengua de los médicos son diferentes a las de los ingenieros, los primeros requieren mayor desarrollo en la comprensión lectora, mientras los segundos de interacción oral.

Cuando van a entrevistas lo que a ellos (alumnos de Ingeniería) más les demandan es pues que se sepan desarrollar en una conversación, no qué tanto entienden en una lectura o que tanto dominan de voz pasiva, de *reported speech* o cosas así no o composiciones de por ejemplo les pedimos ehm escríbele una carta de darle agradecerle algo a alguien. Me dicen es que eso nunca me lo van a pedir ¿no? entonces yo por eso lo veo en lo que ellos necesitan es hablar, mis médicos necesitan leer porque tienen que pasar el examen nacional de residencias y el examen o una buena parte o una parte vienen en comprensión de textos en inglés entonces las ideas son diferentes (Camila: Entrevista, 6:16; 6:20).

Esta reflexión de la profesora es trascendente, ya que muestra sensibilidad hacia las necesidades de los estudiantes, más acorde a la vida real, así como habilidad de su parte de adaptación al contexto.

La validez como concepto técnico no está tan presente en el discurso de los profesores. Si bien, algunos lo mencionan, la mayoría expresa alguna manifestación o aplicación del concepto de manera empírica o de alguno de los tipos de validez, como la validez de

contenido o la autenticidad, los cuales sí tienen mucha mayor presencia que el concepto en sí.

Otro factor de importante presencia en el discurso de los profesores ha sido los aspectos relacionados con la practicidad que se explican a continuación.

4.3.2.3 Practicidad

El código denominado Practicidad hace referencias a los recursos necesarios en el diseño de una evaluación, así como a expresiones que impliquen un uso adecuado de éstos. Este aspecto requiere de amplia consideración en el diseño de exámenes, ya que, por razones institucionales o personales, los recursos como el tiempo y el dinero siempre serán limitantes en la práctica evaluativa. En las entrevistas a profesores se encontraron referencias a este tema.

Dentro de este código se contemplan también las referencias a hacer más fácil para los maestros el trabajo relacionado con los exámenes de lengua, hablando de facilidad en términos de un uso óptimo de recursos. Se desarrolló otro código denominado Facilidad, pero este fue utilizado para referirse al bajo grado de dificultad de un examen.

Nora se refiere a la practicidad como una característica de un examen, particularmente refiriéndose al tiempo como un recurso limitado.

Otra característica de los exámenes creo que debe de ser el tiempo, tiene que limitarse al tiempo de tu clase, si tu clase es de 40 minutos creo que tu examen debería respetar una hora y media no les puedes dar un examen que lo terminen en media hora... (Nora: Entrevista 3:62).

Erika se expresa acerca del tiempo que se requiere para realizar un examen, al responder una pregunta sobre algunas solicitudes de la coordinadora a los profesores.

... o sea yo creo que nos costaba, pero por esta parte de diseñar un examen así yo creo que sí te toma pues mínimo una semana o sea no es algo tan rápido o sea no es de que digas ahorita voy a hacer un examen en dos horas (Erika: Entrevista, 2:44).

Estas reflexiones acerca del tiempo que toma diseñar un examen son evidencia de un aspecto trascendente en el proceso de diseño de instrumentos de evaluación. Sin embargo no es el único. La practicidad también se relaciona con los recursos monetarios, físicos o procesos logísticos y administrativos (Erika: Entrevista, 2:116; Alondra: Entrevista, 6:165)

Para otros profesores, como Graciela, diseñar un examen implica tiempo, trabajo y dinero y se manifiesta inconforme con la retribución que obtiene por este trabajo (Graciela: Entrevista, 8:105), Incluso cuando se le preguntó qué decían sus exámenes de ella como profesora manifestó en primer lugar la practicidad (Graciela: Entrevista, 8:137).

La practicidad es un aspecto que se relaciona con el diseño, la aplicación y la calificación de exámenes. Es esta última la que predomina en el discurso de los profesores, sobre el cual la mayoría comenta que un examen debe ser fácil de calificar y los recursos necesarios para el diseño, entre los que destaca el tiempo.

4.3.2.4 Dificultad

El código Dificultad se refiere a aquellas citas que expresan dificultad en la resolución o diseño de una evaluación.

Profesores como Lorena se expresan sobre la dificultad que representa diseñar instrumentos que puedan medir el aprendizaje y hacerlo objetivamente. En la siguiente cita, Lorena habla de la importancia que tiene que los profesores se capaciten técnicamente

... yo creo que sí es importante porque si no estás diseñando un instrumento que no te sirve para nada, estás cumpliendo con el mero requisito de hacer un examen, pero no estás midiendo de verdad el aprendizaje, que ya de por sí medir el aprendizaje es muy complicado... (Lorena: Entrevista, 1:91).

... y es muy difícil que llegue a ser objetivo y que mida lo que de verdad sabes... (Lorena: Entrevista, 1:259).

Erika comenta las dificultades que en el proceso de diseñar exámenes se tiene con el trabajo colaborativo (Erika: Entrevista, 2:84-85) y también manifiesta la dificultad de calificar ciertos aspectos de los exámenes, como explicó en el momento de hablar sobre la confiabilidad de los exámenes (Erika: Entrevista, 2:120-121).

Otras manifestaciones de dificultad implican las dudas que generan en el profesor los exámenes que han sido difíciles (Camila: Entrevista, 4:48); el balance necesario en el grado de dificultad de los reactivos (Gina: Entrevista, 5:33); el diseño de cierto tipo de tareas como las de comprensión oral (Alondra: Entrevista, 6:109-110) o de comprensión escrita (Genaro: Entrevista, 10:47) o las relacionadas con aspectos culturales (Virginia: Entrevista,

11:57-59). Todos estos aspectos se relacionan directamente con la autenticidad de las tareas en un examen, que a su vez tiene impacto en la validez.

El análisis muestra que los profesores encuentran dificultad en una variedad de aspectos de evaluación, como medir el aprendizaje, el trabajo colaborativo, y calificar, diseñar o darle un mayor grado de validez a algunas tareas en los exámenes.

Algunos profesores parecen utilizar la palabra “complejidad” para referirse a una dificultad.

... normalmente hago reactivos extra por ejemplo calculo cuantos reactivos de cada ejercicio de acuerdo también al nivel de complejidad, o de la tarea que se requiera calculo, por ejemplo, esta este ejercicio no requeriría más de 5 reactivos o más o este si requiere ese es un ejercicio de complejidad alta este si se requiere a lo mejor 5 pero con a cada respuesta se le daría dos aciertos... (Simón: Entrevista, 7:36-37).

... porque trato de poner una variedad de ejercicios que les permitan a ellos tener diferentes niveles de complejidad... (Lorena: Entrevista, 1:145).

Parece ser que cuando los profesores hablan de un “nivel de complejidad” se refieren a un grado de dificultad.

En su discurso, los profesores expresan dificultad para ellos mismos, así como para los alumnos. Cuando se refieren a dificultad para sí mismos, generalmente se refieren a temas relacionados con alguna fase del proceso de evaluación, ya sea diseño o calificación del instrumento, que a su vez se relacionan con el grado de validez, confiabilidad, autenticidad o practicidad. Cuando hacen referencia a los alumnos, es con relación a la resolución del examen, y a la vez implican justicia y retos para ellos.

4.3.2.5 Adaptación

El código Adaptación se refiere a cualquier alusión al uso de materiales o textos existentes en el diseño de una evaluación. Los profesores manifestaron utilizar diversas fuentes y procedimientos para el diseño de sus exámenes. Esta práctica surgió en la fase cuantitativa y se decidió corroborar con las entrevistas.

Algunos, como Camila y Genaro, se expresaron con respecto a la adaptación de materiales existentes desde la pregunta sobre cómo diseñaban sus exámenes.

... tomo de los libros, más bien material de los libros que nos proporcionan para el material de maestros, ciertos reactivos que me pudieran ayudar a conformar cada área del examen... (Camila: Entrevista, 4:21).

... sí pero no trato de tomar todo el examen del material que me dan, sino como conformarlo, hacer un híbrido de lo que hay... (Camila: Entrevista, 4:127).

A algunos profesores se les preguntó de manera directa si sus reactivos eran de diseño propio, los tomaban de algún lado o una combinación de ambos. En los casos de Lorena (Lorena: Entrevista, 1:16 y 1:20-21) y Nadia (Nadia: Entrevista, 1:20-21), comentaron que era este último.

Graciela toma los exámenes de progreso que se incluyen en los libros de texto (Graciela: Entrevista, 8:17-18), en contraste con el inicio de su práctica profesional, cuando adaptaba materiales de otros libros de texto debido a que éstos no incluían los exámenes de práctica (Graciela: Entrevista, 8:88-89). Belinda (Belinda: Entrevista, 9:77-80) y Nadia (Nadia: Entrevista, 11:39-40) expresaron utilizar materiales existentes porque les ahorra tiempo

Genaro manifiesta que ha realizado pocos exámenes, lo que implica que la mayoría son adaptados o tomado de tareas ya existentes. Esto lo manifestó al responder a la preguntaba sobre qué pensaba de sus exámenes:

De los pocos que he hecho porque tampoco he hecho muchos, la mayoría los he tomado de otras partes... (Genaro: Entrevista, 10:112).

Genaro justifica esta práctica alegando validez de las tareas tomadas de otros libros de texto, al responder la pregunta directa sobre si la práctica es válida, aunque en su respuesta inicial no fue así, al momento de contestar hace una reflexión:

Yo creo que le quita validez, no, bueno es que, ni tanto, bueno, yo creo que sería válido más bien, porque si tenemos diferentes libros de texto pero de nuevo los libros están enfocados en un nivel A1 B1 B2 B3 pues se supone que están utilizando el vocabulario las estructuras que se espera que tengan en ese nivel entonces no creo que le quite validez a menos que no lo hayamos visto en clase... (Genaro: Entrevista, 10:118-120).

La justificación de Genaro toca el aspecto de validez, específicamente la de contenido.

La adaptación de materiales existentes para instrumentos de evaluación parece ser una práctica generalizada. El análisis de discurso muestra evidencia de adaptación, principalmente de textos orales y escritos y de reactivos. Entre las razones por las que se realiza esta práctica, los profesores mencionan comodidad, desconfianza en su propio trabajo, aunque la reflexión final de Genaro es acerca de la validez del instrumento resultante de esta práctica.

4.3.2.6 Preparación y conocimiento técnico

En este inciso se analizan dos códigos que se relacionan entre sí, debido a que se refieren a aspectos técnicos específicamente preparación y conocimiento. Aquellas referencias a cursos o preparación técnica específicamente en diseño de evaluaciones o exámenes de lengua han recibido la codificación denominada Preparación técnica. El código Conocimiento técnico se refiere a aquellas implicaturas al conocimiento técnico que se requiere para elaborar o diseñar un examen de lengua.

En las entrevistas se les preguntó a los profesores si habían recibido capacitación específica en cuanto a diseño de exámenes, a lo que la mayoría (Lorena: Entrevista, 1:75; Graciela: Entrevista, 8:99; Genaro: Entrevista, 10:103) respondió afirmativamente, generalmente en los estudios de licenciatura.

Otros profesores, como Erika (Erika: Entrevista, 2:80) y Camila (Camila: Entrevista, 4:138-141), manifestaron haber tomado cursos durante su práctica profesional lo cual coadyuvó a que las profesoras percibieran un cambio en su trabajo como diseñadores de instrumentos de evaluación. Graciela, en contraste con quienes piensan que mejoraron su práctica, considera que el contenido del curso tomado se aleja de la realidad del profesor. Consideró que cuando ella estudiaba los exámenes no tenían la rigurosidad que se pedía en ese curso (Graciela: Entrevista, 8:102).

Entre los profesores existe un reconocimiento de la necesidad de desarrollo académico en cuanto a evaluación y diseño de exámenes

Entonces, sí necesitamos mucha capacitación en la evaluación y mucha más capacitación en diseño de exámenes... (Nora: Entrevista, 3:211).

Entonces, creo que sí, sí es necesario y yo sería partidaria de que sí se tomara, bueno que sí se dieran cursos de formación para elaborar un instrumento... (Camila: Entrevista, 4:82).

De igual manera, Nora (Nora: Entrevista, 3:208), Erika (Erika: Entrevista, 2:90-92), Camila (Camila: Entrevista, 4:79-80) y Nadia (Nadia: Entrevista, 12:73) reconocen la importancia de este la preparación técnica

En el discurso de los profesores puede percibirse reflexión sobre la importancia de preparación en evaluación y la necesidad de ello. Se reconoce el interés por parte de los

directivos de las instituciones donde laboran los profesores. Puede concluirse que a pesar de la importancia que se le da a la evaluación, parece que en la formación básica de los profesores no es un aspecto al que se le da mucha importancia, sino que es en la práctica profesional el momento en que se despierta ese interés por el conocimiento técnico y especializado.

4.3.2.7 Tecnicismo

El código Tecnicismo se utilizó para identificar el uso de términos técnicos relacionados con el diseño de instrumentos de evaluación. Este código tiene relación con la preparación y conocimiento técnico.

El código Tecnicismo se refiere al uso de jerga relacionada al diseño de exámenes, aunque es común en el discurso de los profesores encontrar términos que no corresponden a ésta. Por ejemplo, anteriormente en este análisis, se encontró que los profesores utilizan, en su mayoría, el término “ejercicio” cuando el término “tareas” o “técnicas” es más adecuado para referirse a las actividades a realizar, por parte de los alumnos, en la resolución de un examen de lengua. En este análisis se dará cuenta del uso de terminología adecuada, así como de términos utilizados para compensar la falta de ella.

El uso del término “reactivos” es de los más usados entre los profesores entrevistados. Camila los usa en repetidas ocasiones durante su entrevista, por ejemplo, al momento de describir cómo diseña sus exámenes,

... primero reviso cuáles son los temas que se vieron durante ese periodo, posteriormente veo de esos temas cuales si se podrían a lo mejor evaluar a través de reactivos escritos... (Camila: Entrevista, 4:17-18).

Alondra se refiere a los reactivos de opción múltiple cuando da su opinión sobre este tipo de técnica

Para nosotros como maestros y diseñadores de exámenes es mucho más fácil mucho más factible, ah para calificar no así el diseño porque el diseño involucra saber desde diseño de reactivos ¿no? de reactivos de opción múltiple... (Alondra: Entrevista, 6:47-48).

Es de uso común entre los profesores entrevistados el uso de terminología para referirse al tipo de técnicas o tareas de evaluación, como opción múltiple, *fill in the gaps*,

Por ejemplo, Nora al describir un problema que tuvo en la aplicación de un examen:

Era un *fill in the gaps* y un *true and false* entonces se confundieron mucho, pero yo veía que estaban atentos... (Nora: Entrevista, 3:22).

Junto con Nora, Alondra, una de las participantes que maneja más terminología de diseño de exámenes, al explicar cómo diseña sus exámenes menciona:

Aquí sí hay preguntas abiertas, opción múltiple, este *fill in the gaps* o ¿cómo se le llama? ¿*Cloze*? ajá? (Alondra: Entrevista, 6:84).

Ah, bueno lo que nosotros consideramos es lo que llamamos *face validity*, mmm, ay se me fue la palabra, ay se me olvidó, no sé cómo traducirlo al español, *face validity*, *administration*, *content validity* y y una ¿qué es? *relia*... (Alondra: Entrevista, 6:161-162).

Genaro utiliza los términos validez y confiabilidad (Genaro: Entrevista, 10:75-76) y se refiere a la validez visual (*face validity*) sin mencionar el término técnico. Sin embargo, más adelante se le preguntaría a qué se refería con validez y confiabilidad. En cuanto a validez respondió

Que estén evaluando lo que se tiene que evaluar (Genaro: Entrevista, 10:83).

Lo que es correcto, con base en la literatura especializada. En cambio, al preguntársele sobre confiabilidad su respuesta fue

No lo sé. (Genaro: Entrevista, 10:91).

El escaso uso de tecnicismos o su uso de forma imprecisa es consistente con la falta de preparación técnica de los profesores, la cual pudo observarse en el inciso previo. Sin embargo, estos hechos no demeritan las prácticas evaluativas que de manera empírica han adoptado los profesores en beneficio de los grados de validez y confiabilidad de los instrumentos, y que se han mencionado también en incisos previos. Este fenómeno es un indicio de la existencia de un sistema que se adapta a su contexto, al encontrar una carencia en uno de sus componentes se compensa en otro.

4.3.2.8 Colaboración

El código Colaboración se refiere a las referencias a trabajo realizado junto con colegas profesores. El trabajo colaborativo se da en varios aspectos del proceso de evaluación.

Erika, al responder a la pregunta sobre cómo diseña sus exámenes, inicia describiendo los acuerdos con los profesores de la institución donde labora y cómo dividían el trabajo. En la

segunda cita menciona el proceso de revisión y retroalimentación sobre el examen entre colegas:

...en la última junta acordamos hacer como un *checklist* entonces teníamos que partir de qué contenido íbamos a ver y después procedíamos a dividirlos, éramos seis maestras entonces a cada uno le tocaba hacer un parcial, versión A y versión B... (Erika: Entrevista, 2:25-26).

... y ella (otra maestra) lo tenía que contestar sin ver el *answer key* no, y ya a ver sabes qué Erika, aquí estuvo mal o sabe qué maestra modifique esta parte... (Erika: Entrevista, 2:36).

Alondra también muestra indicios de trabajo de revisión de exámenes entre colegas o en academia (Alondra: Entrevista, 6:110). La profesora explica cómo se dio cuenta de algunas dificultades en el momento de diseñar una tarea para comprensión oral gracias al trabajo de revisión por parte de su academia. Más adelante en la entrevista, explicaría cómo está organizado el trabajo entre los profesores en la institución donde labora

En Ingeniería estamos organizadas de tal manera que nos repartimos quién hace qué examen, pero todas estamos obligadas a revisar todos los exámenes y a contestarlos y hacer observaciones y se forma aparte una comisión revisora... (Alondra: Entrevista, 6:117).

Gina no describe trabajo formal en equipo o acciones como revisiones de exámenes o pilotajes, sino que la relación con los colegas tiene impacto en el trabajo del profesor, gracias a discusiones, opiniones, experiencias, aunque no menciona si existe un proceso formal para esta colaboración o es resultado de las relaciones interpersonales y particulares entre los individuos en la institución donde labora.

... entonces sí, aparte el hecho de tomar el curso como estaban también otros compañeros, entonces el hecho de discutir de cambiar opiniones, de intercambiar también las experiencias eso ayuda bastante, la verdad es que ayuda bastante... (Gina: Entrevista, 5:75).

En contraste con esta declaración, algunos expresaron no estar satisfechos con algunos aspectos del trabajo colaborativo, como relata Nadia lo sucedido después de recibir para revisión el examen de una colega quien solo lo copió de un sitio en internet (Nadia: Entrevista, 12:77). Nadia menciona que el trabajo de un profesor puede afectar a otros, como en este caso que se entregó un examen con fallas que era necesario corregir, lo que requirió del tiempo de los revisores.

La colaboración entre profesores tiene aspectos positivos y negativos, como se muestra en la evidencia presentada, aunque si bien los comentarios con una denotación negativa son importantes para los profesores, las situaciones causantes de esos comentarios son áreas de

oportunidad que bien pudiesen solventarse con mejor organización administrativa o con desarrollo de la literacidad en evaluación de lenguas.

4.3.2.9 Calificación

El código Calificación son las referencias al proceso de calificación (*scoring*) de los instrumentos de evaluación.

Para los profesores es importante que el proceso de calificación sea práctico. Así lo expresa Lorena al explicar que es necesario delimitar el contenido del examen.

... creo que me pueden salir como cuatrocientos reactivos y no, no sería práctico, para calificar...
(Lorena: Entrevista, 1:39).

En el caso de Alondra, es la institución quien ha establecido una manera de evaluar que da practicidad al instrumento a la hora de calificar.

... En todas las materias son de opción múltiple porque obedece a que ellos manejan un sistema de lector óptico de evaluación, o sea nosotros como maestros no calificamos, existe un departamento de evaluación como tal, allá, y entonces todos los exámenes se entregan, las respuestas se entregan con papeleta por eso obedece a que sean de opción múltiple (Alondra: Entrevista, 6:41-42).

Graciela comenta que ella como docente es práctica, y menciona las repercusiones de las calificaciones en su persona, como estrés (Graciela: Entrevista, 8:106-107). Graciela agrega que el tiempo dedicado a calificar exámenes no es pagado, por lo que trata de reducirlo con tareas prácticas de calificar:

Para Virginia, el proceso de calificación le ayuda a obtener información de los estudiantes. Esto lo manifiesta en el momento en que responde a la pregunta qué es lo que dicen sus exámenes acerca de sus alumnos:

A mí me informan si el tema ya está dominado o hace falta retroalimentación sobre todo cuando voy calificando ¡Ah ya lo domina! ya tiene su vocabulario ya es variado, me falta trabajar tal tema o no se... (Virginia: Entrevista, 11:152).

La dimensión técnica se relaciona con aspectos técnicos del diseño de exámenes. El discurso de los profesores muestra evidencia del reconocimiento de la importancia de la técnica por parte de los participantes, quienes a su vez demuestran necesidad de desarrollo en el área. Los aspectos técnicos imperantes son los que se presentan en la literatura: validez, confiabilidad y practicidad y que juegan un papel central en la estructura de esta

dimensión. El resto de los códigos de la dimensión se desarrollan en su mayoría alrededor de estos tres, ya sea contribuyendo a su logro o con alguna relación con ellos. Estas conexiones entre códigos sugieren la presencia de un sistema complejo, que a pesar de la falta de conocimiento técnico de los participantes compensa para el logro de alguno de los tres aspectos técnicos principales, gracias a las situaciones particulares y contextuales de los participantes.

Las conexiones mencionadas no solamente se dan al interior de la dimensión técnica, sino que presentan fuertes relaciones con otras dimensiones, destacando la dimensión ética, que se presenta a continuación.

4.3.3 La dimensión ética en el discurso de los profesores.

La dimensión ética agrupa aquellos códigos que implican acciones o pensamientos que representan bienestar. Está conformada por 69 códigos que surgieron del análisis del discurso, entre los que destacan por su frecuencia: Alteridad, Cuidado del alumno, Exclusión, Inclusión, Reflexión, Impacto, Diálogo, Autoevaluación, Juicio al alumno, Valoración, Obligación, Justicia e Importancia.

En este inciso se presentan los códigos mencionados y otros de menor presencia en el discurso de los profesores. Los códigos de la dimensión ética se presentan en dos grupos: la subdimensión interpersonal y la valorativa. La subdimensión interpersonal se refiere a los códigos que evidencian la estructura y mecanismos de las relaciones entre las personas participantes en el proceso. La subdimensión valorativa se refiere a aquellos códigos que presentan juicios con respecto a las personas que participan en la práctica evaluativa de la enseñanza de lenguas o a los instrumentos y procesos con que se llevan a cabo.

4.3.3.1 Subdimensión interpersonal

4.3.3.1.1 Alteridad

El código Alteridad, es junto con Confiabilidad, de la dimensión técnica, el de mayor presencia en el discurso de los profesores. Se utilizó este código para las citas que hagan referencia o reconocimiento a la existencia de un tercero, generalmente un alumno. Cabe mencionar que la alteridad no implica un juicio o valoración de otra persona, sino que es un

posicionamiento del hablante en el lugar de otro participante en el proceso evaluativo, para llegar a una conclusión o elaborar un argumento ante la pregunta del investigador.

Estos son algunos ejemplos del posicionamiento mencionado renglones arriba.

En la siguiente cita, Lorena se encuentra hablando de lo difícil que es diseñar instrumentos válidos y ejemplifica con un posicionamiento en el lugar de un alumno

...un alumno puede aprenderse las cosas de memoria, y ponértelas en el examen y sacar diez...
(Lorena: Entrevista, 1:93).

La profesora no habla de un alumno o con referencia a una experiencia en particular, pareciera estar haciendo una generalización con este ejemplo. No se encuentra evidencia en su discurso de conocimiento de que lo que describe lo sepa de voz de algún alumno, sino que se ubica en su lugar, al parecer, haciendo una suposición de lo que podría estar sucediendo.

En ocasiones, el profesor hace referencia a los alumnos y sus acciones gracias a evidencia obtenida por experiencia o por observación, como Nora quien explica cómo estructura su examen y que los alumnos no necesariamente contestan el examen en ese orden (Nora: Entrevista, 3:60).

Los participantes también utilizan la experiencia propia como estudiantes para ejemplificar lo que sucede con los alumnos en las evaluaciones. Camila explica que los exámenes deben ser claros porque de lo contrario desgastan al alumno (Camila: Entrevista, 4:43) poniéndose en el lugar del alumno y tomando su experiencia como tal.

Dentro de este código también se consideraron aquellas citas que hacen referencia a la conciencia de la necesidad de entender o ponerse en el lugar del alumno, como lo hace Lorena en la siguiente cita, donde la profesora se encuentra hablando de la necesidad de tener un perfil adecuado a la labor del docente de lenguas:

...pero quien no tenga el perfil pues mmm como, como que le cuesta más trabajo entender ese proceso, como que le cuesta más trabajo llegar a sus alumnos... (Lorena: Entrevista, 1:243).

La alteridad también se observa en el discurso cuando los profesores buscan encontrar causas por las que los alumnos tienen ciertos resultados. Por ejemplo, Lorena comenta que

la evaluación es abstracta y compleja y que en ocasiones no es posible lograr la validez que se busca en los instrumentos:

... y aunque el alumno sí sabe, en ese momento a lo mejor o tenía un problema en su casa o se distrajo o se sentía mal o no durmió y todo eso afecta también en la calificación, son demasiados elementos que se tienen que considerar... (Lorena: Entrevista, 1:254).

La alteridad se encontró en el discurso de todos los profesores participantes en el estudio y presenta relaciones con otros 99 códigos. Entre los más sobresalientes, debido al número de co-ocurrencias en citas de las entrevistas, se encuentran Juicio al alumno, Cuidado del alumno, Confiabilidad y Habilidades.

Alteridad y Juicio al alumno

En ocasiones la alteridad se ve acompañada de un juicio por parte del participante hacia un tercero. Por ejemplo, en la siguiente cita, Nora explica qué es lo que dicen sus exámenes de sus alumnos, algunos alumnos tienen un nivel más alto que lo que muestra el resultado de su examen, pero por razones, supuestas por ella, no lo logran:

Tengo alumnos que son más altos que su examen, pero de esa manera, tienes mucha fluidez, hablas muy bien, pero no te gusta leer, o no pones atención a la gramática entonces la necesitas. (Nora: Entrevista, 3:143).

Graciela comenta que para ella es importante la honestidad y contrasta con que algunos alumnos hacen trampa en sus exámenes:

... creo que no nada más es que sepan las cosas, sino que sean personas honestas, sobre todo pues por el momento que estamos viviendo, por el país en el que vivimos, entonces me ha tocado que hay alumnos que son, algunos alumnos que, pues son así medio, pues que les gusta copiar ¿no? (Graciela: Entrevista, 8:71-72).

El juicio sobre el alumno es un código que se ha ubicado dentro de la subdimensión valorativa, lo que es evidencia de interconexión entre las subdimensiones de carácter ético identificadas en el análisis.

Alteridad y Cuidado del alumno

Los participantes hablan en repetidas ocasiones de los alumnos y en particular sobre el cuidado que tienen de ellos, de su aprendizaje, de su bienestar. Este cuidado, generalmente,

no se expresa de manera explícita, sino que lo hacen de diversas formas, como se verá en las siguientes citas.

Lorena expresa cuidado del alumno y de su aprendizaje:

... ya sabes que es típico que en la clase les preguntas (a los alumnos) si entendieron y dicen que sí, haces ejercicios y los hacen bien pero cuando pasan una semana dos semanas o tres o llega el día del examen ya no se acuerdan, entonces este creo que sí es importante verificar si de verdad están aprendiendo eso a lo largo del tiempo... (Lorena: Entrevista, 1:267-268).

Nora explica que tiene cuidado con los contenidos de los exámenes por los alumnos.

... como que (a los alumnos) les angustia la gramática, entonces sí soy muy cuidadosa en cuestiones de los temas gramaticales que incluyo... (Nora: Entrevista, 3:91).

Simón considera que el vocabulario es una dificultad para los alumnos por lo que dedica tiempo para trabajar en clase y ayudarles en los exámenes.

... sí enfatizo mucho y a lo mejor no en el examen, pero también en la clase enfatizo mucho con vocabulario, porque sé qué es lo que a veces cuesta trabajo para los alumnos o que esto les pueda ayudar en los exámenes... (Simón: Entrevista, 7:64-65).

Si bien la alteridad representa el reconocimiento de la existencia del otro, la mayoría de los extractos que funcionan como evidencia de este código tienen una relación con aspectos técnicos como confiabilidad y validez. Asimismo, en la descripción del discurso de los profesores la alteridad, que se muestra principalmente en relación con el alumno, ya sea de manera valorativa, como en las conexiones con el código Juicio sobre el alumno, o de manera interpersonal, como con el código Cuidado del alumno, el cual se analiza a continuación.

4.3.3.1.2 Cuidado del alumno

El código Cuidado del alumno representa aquellas citas que se refieren a una acción o pensamiento en pro del bienestar del alumno. En el inciso anterior se mostraron algunos ejemplos de este código que se relaciona intrínsecamente con el código Alteridad.

Estas son algunas manifestaciones del cuidado del alumno. En el momento de justificar el uso de exámenes para evaluar, Lorena (Lorena: Entrevista, 1:5) y Simón (Simón: Entrevista, 7:99-100) expresan cuidado del alumno al manifestar interés por el avance en su aprendizaje. Erika explica que, al revisar los exámenes de los colegas, se requiere hacerlo con objetividad, entre otras cosas, por los alumnos (Erika: Entrevista, 2:86). Belinda

(Belinda: Entrevista, 9:254) y Lorena (Lorena: Entrevista, 1:207) explícitamente expresan la importancia del impacto del examen en los alumnos. La importancia de la preparación del profesor en referencia al cuidado de los alumnos fue expresado por Lorena (Lorena: Entrevista, 1:234), Erika (Erika: Entrevista, 2:93) y Camila (Camila: Entrevista, 4:109-110).

Algunos profesores expusieron una serie de medidas que toman durante el diseño del examen y las relacionan directamente con sus alumnos, por ejemplo: Adaptar las tareas del examen por el alumno (Lorena: Entrevista, 1:16), incluir las rúbricas de habilidades productivas en el examen (Lorena: Entrevista, 1:45), elegir diferentes tipos de tareas, (Lorena: Entrevista, 1:145), compensar en la calificación por tareas difíciles o injustas (Camila: Entrevista, 4:151-152), familiarizar a los alumnos con las tareas a realizar en el examen, previo a este (Simón: Entrevista, 7:34) e incluso utilizar pocas hojas de papel de manera que el alumno no perciba que es muy largo y se dé un impacto negativo (Nora: Entrevista, 3:190-191).

Por otra parte, Graciela expresa cuidado del alumno en su discurso al señalar que además del aprendizaje, el desarrollo personal es importante:

... porque no solamente van a aprender cosas, creo que también tienen que formarse como personas... (Graciela: Entrevista, 8:9).

... para que pues aparte de... aprender conocimiento (*sic*) tengan también crecimiento como personas (Graciela: Entrevista, 8:11).

El cuidado del alumno por parte del profesor se expresa de varias formas: Revisar su avance, darles algo, mejorar su aprendizaje, compensar cuando se dan injusticias, considerar el impacto o cómo les va a afectar alguna acción, no perjudicar, que no se estresen, cómo ayudar, contribuir en su crecimiento, formación y aprendizaje.

De igual manera, se mencionan algunos aspectos técnicos, específicamente aquellos relacionados con tareas, rúbricas, y longitud de exámenes y cómo los alumnos se podrían ver afectados por las decisiones que toman los profesores en este sentido.

En esta ética interpersonal, los profesores nos solo reconocen en su discurso la existencia del Otro, el alumno, sino que muestran un sentido de responsabilidad, de cuidado hacia él,

y también la expresan en un reconocimiento como parte, o no, de una relación, comunidad o actividad, que se han categorizado como exclusión e inclusión.

4.3.3.1.3 Exclusión e inclusión

El código Exclusión se definió por sus características discursivas, se describió como aquellas citas que especifican la primera persona del singular (yo) o la tercera persona del singular o plural (él, ella, ellos) como sujeto de un enunciado. Se consideró que esta manera de expresarse de los participantes tiene una implicatura, el excluir de la acción a otros o excluirse a sí mismos de las acciones de otros. En este inciso se presentan ejemplos de ambos casos.

Exclusión y el proceso de diseño de exámenes

La exclusión fue un código recurrente, especialmente cuando los participantes explicaban el proceso en que diseñan los exámenes. Como es el caso de Nora en la siguiente cita:

... entonces veo esos temas, después trato de buscar ejercicios que yo tenga que ya haya ocupado antes, que sean de ese tema, entonces tengo un pequeño compendio de ejercicios gramaticales que a veces no utilizo en clase con mis chicos y entonces los dejo ahí y a veces los utilizo y los pongo en los exámenes de gramática... (Nora: Entrevista, 3:16).

Lorena explica cómo diseña los exámenes y se expresa sobre sus preferencias acerca de la estructura del examen

... me fijo mucho en, en estructurar bien el examen no es nada más como encontré unos ejercicios y los puse sino más bien qué quiero que hagan aquí... (Lorena: Entrevista, 1:210-211).

Alondra menciona que ella en el diseño de exámenes busca evaluar ciertos aspectos de la lengua y hace el contraste con lo que hacen otros profesores:

... al menos yo procuro eso ¿no? que no nada más se refleje la memorización y que (El alumno) sabe la estructura (gramatical) sino que aparte discrimina entre cuando usar uno de los otros dos, pero si hay unos maestros que si lo hacen así... (Alondra: Entrevista, 6:37-38).

De esta manera Nora, Alondra y Lorena expresan por medio del Yo ser la responsables, quienes han decidido, convirtiéndose además en la ejecutoras de las prácticas de diseño descritas, independientemente de su contexto.

Exclusión y dificultades en el diseño de exámenes

Por otro lado y en relación al proceso, algunos profesores como Simón y Genaro manifestaron dificultades al momento de diseñar exámenes:

.... yo cuando diseño exámenes siempre siempre tengo en consideración muchos, digo para ser un examen a mí me cuesta trabajo, porque he aprendido que tiene que ser confiable tiene que tener (*sic*) validez que entonces trato de equilibrar todas estas partes en mis exámenes... (Simón: Entrevista, 7:8).

... la parte más difícil siempre ha sido seleccionar algún texto porque no se a mí en lo personal no se me da hacer las preguntas de comprensión... (Genaro: Entrevista, 10:47).

Estos profesores expresan específicamente en cuanto a su persona, que son ellos exclusivamente, quien presenta dificultades.

Sobre la pregunta si se habían tenido dudas sobre los exámenes, la mayoría manifestaron afirmativamente, en contraste Belinda expresó:

Cuando los he hecho, no, (Belinda: Entrevista, 9:98).

Aquí la profesora expresa que no hay dudas o dificultades en el examen cuando ella misma lo realiza, lo que implica que sí ha sido así cuando colabora con otros colegas. Es decir, la exclusión de estos en la elaboración de los exámenes también elimina las dificultades.

Exclusión y preparación técnica

Lo mismo se da en referencia a la preparación técnica requerida para las prácticas mencionadas. Alondra conservaría la misma postura con respecto a sus colegas más adelante en la entrevista, donde habla sobre la preparación técnica para diseñar instrumentos de evaluación:

...de hecho, se nos ha dado capacitación que igual no todos los maestros toman ¿ajá? pero sí se ha dado capacitación para el diseño de exámenes de opción múltiple... (Alondra: Entrevista, 6:49).

Genaro también opina sobre la capacitación técnica, pero él lo hace refiriéndose a los profesores en tercera persona, excluyéndose.

... entonces creo que sí, los profesores deberían tener esta parte de así se hace un examen, para hacer algo más uniforme... (Genaro: Entrevista, 10:111).

Tanto Genaro como Alondra plantean la necesidad de capacitación técnica de los otros profesores, aunque no se incluyen a sí mismos en tales afirmaciones.

Exclusión y el alumnado

La exclusión se da en suposiciones que hacen los profesores sobre los alumnos. En este caso sobre sus prácticas de aprendizaje, como se muestra en la siguiente cita:

...porque yo estoy asumiendo que mis alumnos están leyendo, y a veces no... (Lorena: Entrevista, 1:148).

Por ejemplo, Erika comenta su inconformidad con los reactivos de opción múltiple por la manera en que, ella supone, los alumnos contestan este tipo de reactivos:

... porque a veces siento que ellos (los alumnos) nada más ponen al azar... (Erika: Entrevista, 2:70).

... por ejemplo había un examen de ubicación ese me llama mucho la atención porque uno se esfuerza en que realmente reflejen no, es muy importante en qué nivel van a llegar los chicos y para ellos a veces ni siquiera los leen los contestan al azar... (Erika: Entrevista, 2:101).

La exclusión del alumnado se da con referencia a ciertas prácticas de aprendizaje cuestionables o a la ausencia de prácticas eficientes, como contestar los exámenes “al azar”, “aprender de memoria” o la falta de lectura.

Aunado a la exclusión del profesor por desacuerdo con las prácticas realizadas por otros integrantes del proceso evaluativo. De igual manera también expresan exclusión al relacionar su comentario con algún objeto de evaluación como la gramática:

Exclusión y evaluación de la gramática

Erika narra que la coordinadora de la institución donde trabaja les pedía hacer exámenes con un contexto y con relación entre las diferentes tareas. La profesora comenta ella prefería hacer exámenes de gramática:

... por ejemplo algo muy importante que a mí me costaba al principio trabajo que yo quería hacer de cada tema de gramática... (Erika: Entrevista, 2:51).

Más adelante en la conversación, Erika confirma su gusto por evaluar el aspecto gramatical:

... yo creo que daba mucho énfasis a eso no, a poder ver ah pues sí vimos... el presente, el pasado creo que a nosotros como maestros nos gusta... (Erika: Entrevista, 2:51).

Puede observarse que Erika disiente con su superior, la coordinadora, con respecto al objeto de evaluación.

En seguida en la entrevista, Virginia explicaría por qué evalúa con mayor peso en la gramática. Ella trabaja en la Facultad de Lenguas donde los alumnos estudian la licenciatura en lenguas.

... pero esto tiene una razón de ser, para mí, están en la Facultad de Lenguas y creo que es necesario que la gramática este bien afianzada, más que las otras habilidades... (Virginia: Entrevista, 11:23).

Parece ser que la gramática como un objeto de evaluación causa disenso entre la comunidad académica, ya que varios de los participantes muestran indicadores de exclusión al referirse a este aspecto.

Exclusión y particularidades en la práctica evaluativa

Algunos participantes se expresan de manera excluyente para precisar un aspecto muy particular de su práctica evaluativa, como es el caso de Virginia. La profesora explica que en sus exámenes se enfoca en las excepciones, más que en las reglas gramaticales generales.

... yo evalué mucho las excepciones que se aprendan más la excepción a la regla... Pues porque es las excepciones es lo más difícil... entonces, quiero ver si sí aprendieron la irregularidad, no la regularidad... (Virginia: Entrevista, 11:34-37).

De igual manera, comenta otra práctica muy particular, la reiteración de contenidos de los exámenes:

... yo les hago quiz de gramática eso sí son exclusivos de gramática y vocabulario, entonces en los quiz aparecen los temas gramaticales y si un tema gramatical no es dominado vuelve a aparecer en el siguiente quiz, no desaparece hasta que no esté dominado. (Virginia: Entrevista, 11:154).

... también refleja un, un cierto, mira la gente que llega a mi salón de clase debe dominarme el tema... (Virginia: Entrevista, 11:184).

Estas prácticas descritas por la profesora no son comunes y al parecer está consciente de ello porque manifiesta su exclusión del resto de los profesores al enfatizar con un Yo que ella sí las realiza.

Exclusión y centralidad del profesor

Se encontró en el análisis de discurso que una forma de exclusión es aquella en que los profesores manifiestan indirectamente ser elementos centrales o se otorgan importancia con respecto a los hechos que explican.

En la siguiente cita, se ejemplifica como Nora implica que los alumnos tienen una obligación con ella:

... a veces (los alumnos) pueden ser muy buenos en gramática o en comprensión oral o escrita pero también necesito asegurarme que de alguna manera ellos me demuestren el nivel de lengua... (Nora: Entrevista, 3:41).

Nora manifiesta que necesita que los alumnos le demuestren a ella su nivel de lengua, excluyendo a cualquier otro participante del proceso, como si el objetivo de la evaluación fuese precisamente la demostración a su persona por parte de los alumnos. Nora, utiliza el pronombre “me”, de manera que parece indicar una obligación o responsabilidad de los alumnos hacia ellas

Graciela se muestra muy exclusiva en cuanto a las decisiones que toma en el diseño de sus exámenes:

... entonces tomo lo que estuve viendo en clase, lo que considero que es necesario que los chicos sepan... (Graciela: Entrevista, 8:14-15).

... entonces yo de esos selecciono los que yo considero que tenga que incluir en el examen... (Graciela: Entrevista, 8:18).

De esta manera, Graciela también parece auto otorgarse un papel central en el proceso de evaluación al expresar que es ella quien decide qué es necesario para los alumnos y el contenido del examen.

El uso de los pronombres *me* y *yo* denotan exclusión, limitando ciertos aspectos del proceso evaluativo a su persona, excluyendo a los demás participantes, como pueden ser los alumnos o personal directivo de las instituciones donde laboran. Estas expresiones de exclusión aluden temas diversos como el desempeño de los alumnos y el contenido de cursos o de exámenes.

Los profesores manifiestan exclusión en referencia a varios aspectos ejemplificados en este inciso como el proceso de diseño de exámenes y dificultades en el mismo, la preparación técnica, la evaluación de la gramática y la centralidad del profesor en el proceso, así como hacia algunas prácticas de aprendizaje del alumnado.

En contraste a las expresiones de exclusión, el profesor muestra inclusión en diversos aspectos como se muestra en seguida.

Inclusión

En ocasiones los maestros expresan ser parte de una comunidad con los alumnos o con otros colegas, para señalar este hecho se propuso el código Inclusión que se refiere a aquellas citas donde se utiliza la primera persona del plural para denotar pertenencia a un grupo o comunidad y que son evidencias de inclusión, manifestaciones en las que los profesores comparten algún aspecto del proceso de evaluación con otros participantes.

Inclusión, profesores-alumnado

La primera forma de inclusión que surgió en las entrevistas fue la de los profesores a sus alumnos. En la siguiente cita, Lorena, al describir el proceso en que diseña sus exámenes, menciona:

Bueno, primero identifico cuáles son los elementos conceptos funciones lingüísticas que estuvimos revisando... (Lorena: Entrevista, 1:6).

Esta inclusión se da para describir una acción conjunta en el salón de clase como “revisar funciones lingüísticas”. Por “revisar” la profesora se refiere probablemente al proceso de enseñanza-aprendizaje.

Graciela relata que en una ocasión tuvo dudas con respecto a un examen, debido a que no se practicó en clase la tarea que tendría que realizarse en el examen.

... porque una vez me pasó que no vimos el formato de ese ejercicio y les fue muy mal entonces dije pues creo que ahí metí la pata yo, porque no tuvimos práctica de ese ejercicio en particular entonces por eso creo que les fue tan mal... (Graciela: Entrevista, 8:59).

Esta cita anterior de Graciela muestra diferentes posicionamientos de la profesora. Es inclusiva con los alumnos cuando se refiere a lo estudiado durante el curso, “no vimos el formato de ese ejercicio” o “no tuvimos práctica”. Por otro lado, es excluyente en algunos segmentos de su discurso, por ejemplo, sobre el resultado del examen por parte de los alumnos “les fue muy mal”; al tomar responsabilidad sobre ese mismo hecho, “ahí metí la pata yo”.

Inclusión, profesores como comunidad

Los participantes expresan constantemente su pertenencia a una comunidad o gremio de profesores.

Erika explica sobre la capacitación técnica que recibieron los profesores en la institución donde labora

Sí, en enero nos dieron un curso de aquí de la fac y donde sí nos platicaban mucho sobre la importancia de incluir las rúbricas y ciertas partes del de los exámenes... (Erika: Entrevista, 2:80).

Sobre el mismo tema, capacitación técnica, Genaro y Nadia se incluyen en la comunidad de profesores, al explicar la razón por la que es importante tener esa capacitación:

Sí, sí porque si no, de nuevo ¿cómo estamos evaluando al alumno? Lo estamos evaluando como nosotros creemos... (Genaro: Entrevista, 10:107).

... pero sí siento que es bien importante que constantemente los profesores nos actualicemos para ver qué otras técnicas, qué otros ejercicios, cómo puedo evaluar a mi alumno sin que sea propiamente un examen, porque hay otras formas de hacerlo... (Nadia: Entrevista, 12:84).

Los profesores se expresan como parte de una comunidad sobre temas diversos como la capacitación técnica, aspectos de diseño de instrumentos, trabajo colaborativo, políticas institucionales, estrategias de aprendizaje de los alumnos o el poder que tienen los profesores al momento de evaluar. Parece no existir patrón, por lo que puede deducirse que las diferencias individuales de los profesores son las que los llevan a construir lingüísticamente esta comunidad en la práctica evaluativa.

Inclusión como generalización

La inclusión también fue utilizada por los participantes a manera de generalización. Por ejemplo, en la siguiente cita, Nora da sus conclusiones sobre el tema:

Yo creo que tanto la evaluación como el diseño de exámenes son dos temas súper importantes en la práctica y en la formación docente. Muy importantes, o sea, en la formación docente y en la práctica porque a veces nos formamos como docentes de lengua y sabemos porque leímos, porque nos lo enseñaron y entonces a lo mejor tenemos como vagamente como la idea de cómo se evalúa y de cómo se hace un examen... (Nora: Entrevista, 3:199-201).

Nora es Licenciada en lenguas por lo que parece estar refiriéndose a ella y a la comunidad de egresados de la Universidad.

En la siguiente cita, Camila comenta sobre la importancia de la capacitación técnica para los profesores

... porque a lo mejor tenemos el conocimiento y podemos transmitir los temas, pero eso no nos garantiza que sepamos hacer un buen instrumento... (Camila: Entrevista, 4:81).

Como una conclusión a la entrevista, Simón reflexiona sobre la importancia del diseño de exámenes para los profesionales de la enseñanza de lenguas y se expresa en la primera persona del plural y parece estar haciendo una generalización

Creo que algo que podría decir a manera de conclusión es que, que sí debemos o que sí es necesario... prestar mucha atención a al desarrollo de los exámenes y en lo que se pueda, en lo que nos concierne quizá tratar de buscar preparación, que a veces no siempre tenemos acceso, que no siempre nos damos el tiempo, sobre todo eso, digo también debido a nuestras responsabilidades... (Simón: Entrevista, 7:173-174).

Estas generalizaciones parecen describir un perfil o a su propia comunidad de profesores con respecto a la evaluación. Los participantes presentan un profesor con una vaga, superficial idea de cómo evaluar, sin posibilidades de garantizar una “buena” evaluación.

Inclusión institucional

Virginia hace un comentario que deja ver como ella se siente parte de la institución donde labora. Previo a la siguiente cita, sugiere que los profesores de lengua inglesa y lengua francesa deberían trabajar en conjunto. A ello agrega:

... porque finalmente somos una facultad de Lenguas y finalmente existe un cuadro europeo de referencia para las lenguas no es inglés y francés o sea somos Lenguas e incluso español porque aquí llevan la materia enseñanza de español para extranjero... (Virginia: Entrevista, 11:191).

Virginia expresa identidad institucional, la cual comparte con sus colegas y alumnos, pluralizando y manifestando comunidad.

Los códigos *exclusión* e *inclusión* muestran una extrema complejidad, que es consecuencia del gran número de códigos relacionados. Estos dos códigos son evidencia de la percepción del profesor de su relación con otros participantes del proceso evaluativo, siendo precisos, los alumnos, otros profesores y las instituciones educativas. Esta percepción es variable y es excluyente o incluyente dependiendo de cómo percibe el profesor la vinculación entre él o ella y los otros participantes en relación a un tercer factor o código, como se ha mostrado en este apartado, y entre los cuáles se encuentran el proceso de diseño de exámenes, la capacitación técnica, el trabajo colaborativo, prácticas de aprendizaje, entre varios otros, enunciados arriba.

4.3.3.1.4 Reflexión

El código Reflexión fue considerado dentro de la dimensión ética, identifica aquellas citas que hacen referencia a un pensamiento, diálogo interno, consideración o meditación de parte del hablante. Generalmente implica evidencia en el discurso de tener conciencia del acto de reflexionar o alguna de las acciones mencionadas. Fue común encontrar el verbo “decir” o las frases “darse cuenta” o “ponerse a pensar”. En el análisis se encontró evidencia de reflexión acerca de temas como las prácticas evaluativas, del impacto que pueden tener los instrumentos en los alumnos y sobre los resultados obtenidos por medio del examen. A continuación se da cuenta de ello.

Reflexión sobre las prácticas evaluativas

Los participantes mostraron reflexión para cuestionar sus propias prácticas evaluativas. Los cuestionamientos se dieron específicamente sobre temas como el uso de reactivos existentes en un nuevo instrumento. Es el caso de Camila, quien comenta que en ocasiones utiliza reactivos ya existentes, aunque en ocasiones elabora reactivos propios, de diseño original.

... a veces los incluyo, sí trato de incluirlos (reactivos existentes), pero la verdad a veces digo será que sí se podrá hacer, como que normalmente digo funcionará, no funcionará, pues lo pongo a ver si funciona, me arriesgo a hacerlo... (Camila: Entrevista, 4:31).

En el caso de Gina, hizo una reflexión acerca de su práctica durante un curso para elaboración de exámenes, lo que le hizo detectar áreas de oportunidad:

... entonces el tomar este curso pues me di cuenta de fallitas, ¿no? o sea cuando veía las cosas y lo discutíamos, yo decía ay hójole y veía mis exámenes y decía ay no por qué hice esto, ¿no? y sí tiene toda la razón. (Gina: Entrevista, 5:66-67).

Simón hace explícita una reflexión en su discurso sobre los reactivos que elabora:

... entonces este pues ya al final hago la reflexión y digo no, este tipo de ejercicio mejor no lo ocupo, o lo ocupo en otro nivel, o lo ocupo, no se deberíamos ver esto quizá o debí haberles dado más contexto... (Simón: Entrevista, 7:72).

Otros reflexiones relacionadas con la práctica evaluativa fueron sobre el pilotaje de instrumentos (Belinda: Entrevista, 9:82-83), dudas sobre la propia práctica (Virginia: Entrevista, 11:123-124), mejora en el diseño de los instrumentos gracias a la experiencia (Nadia: Entrevista, 12:66; y Erika: Entrevista, 2:81) y sobre la necesidad de utilizar

exámenes específicamente como instrumento de evaluación (Nora: Entrevista, 3:154). Estas citas revelan que la reflexión de los profesores más común relacionada con su práctica evaluativa es como ésta ha mejorado de sus inicios profesionales a la fecha.

Reflexión sobre el impacto del examen

Los profesores también reflexionan acerca del impacto que pueden tener sus instrumentos en los alumnos. Camila narra una ocasión en que mientras daba clases en la facultad de Medicina, un alumno estaba re-cursando la materia de lengua inglesa y reprobarla significaría baja definitiva del sistema universitario.

... entonces ese momento fue crucial, porque dije...en el examen se decidirá si el alumno se queda o se va, me dio tanto nervio, porque dije h́jole como nada más este examen va a decidir toda su vida académica... (Camila: Entrevista, 4:93-94).

Al parecer, este hecho tuvo impacto en Camila, como se percibe en la siguiente cita de su entrevista:

... y ese hecho me puso a pensar en qué responsabilidad tan grande tengo yo como docente yo como para marcar la vida del alumno... (Camila: Entrevista, 4:97).

En esta cita anterior, Camila expresa haber reflexionado sobre la responsabilidad que tenía ella como profesional de docencia de lengua y su evaluación.

Gina también se manifiesta en este sentido. Explica que los alumnos sufren estrés en las evaluaciones (Gina: Entrevista, 5:91-92) y comentó que es una profesora exigente y que tiene que responder a esta exigencia con sus alumnos, en este caso con un instrumento de evaluación de buena calidad.

El hecho de mostrar reflexión acerca de cómo su trabajo puede afectar, o afecta, a un tercero refleja cuidado, atención, importancia de éste y que a su vez tiene un efecto positivo en la práctica profesional.

De igual manera, los profesores mostraron reflexión ante los resultados que obtienen sus alumnos en las evaluaciones y decisiones que tomaron ante ciertas situaciones, como es el caso de Simón (Entrevista, 7:70), quien comentó que el examen no le dio los resultados que esperaba, o de Belinda (Entrevista, 9:116) quien decidió eliminar una tarea porque confundía a los alumnos; de Genaro (Entrevista, 10:149), quien dudaba de un examen por

su duración; o Erika quien también expresó reflexión acerca del nivel de competencia que estaba evaluando.

Reflexión ante los resultados de un examen

Erika también explica que, al momento de diseñar un examen, el profesor da por sentado muchas cosas, sin embargo, llega un momento de reflexión cuando se aplica el examen.

... pero ya cuando se los das te das cuenta de que no, que a lo mejor tu reacción no fue pues la mejor o tu instrucción no era tan clara. (Erika: Entrevista, 2:81).

En las citas mostradas, los profesores expresan cómo la evidencia del desempeño de los alumnos en una evaluación causó una reflexión sobre su propia práctica, generalmente con un efecto positivo, guiando al docente a la conciencia sobre la necesidad de mejorar dicha práctica.

Hasta el momento se ha presentado evidencia de reflexión manifiesta de parte de los participantes en el estudio. Sin embargo, también en el momento de contestar las preguntas durante las entrevistas se generaba reflexión en los temas tratados.

Reflexión al momento de la entrevista

Durante la entrevista hubo varios momentos en que el tema hacía reflexionar a los profesores.

Por ejemplo, Graciela relata que a sus alumnos les decía que el examen sería corto y en realidad era un examen largo. Al preguntarle la razón de dicha acción, Graciela respondió:

No sé. Ahora que lo pienso no sé, creo que soy mala. Hablando de justicia y mira. Ay, no sé. (Graciela: Entrevista, 8:127).

Belinda comentó sobre una reflexión a la que llegó después de transcurrida la entrevista:

... sí muy interesante, porque ahorita haciendo la entrevista me pongo a pensar en el objetivo, en el efecto que tienen mis exámenes, ¿no?... (Belinda: Entrevista, 9:247).

Genaro compartió que sus exámenes asemejaban mucho a los exámenes de certificación de Cambridge Assessment.

... para *writing* aquí también es un error y lo estoy reconociendo ahorita, es una epifanía, que este, que siempre los hago como más enfocados, de nuevo, a los formatos de Cambridge... (Genaro: Entrevista, 10:65).

Los temas que generaron reflexión en los docentes en el momento de contestar las preguntas de la entrevista fueron sus prácticas evaluativas, en el caso de Graciela y Belinda cuestionando sus propias acciones y en el de Genaro dándose cuenta de un patrón en ellas.

La evidencia muestra que la reflexión tiende a generar consecuencias positivas en la práctica docente y evaluativa. Los profesores manifestaron haber reflexionado sobre dificultades en sus inicios como docentes, específicamente en el área de evaluación y esto ha repercutido en una mejora de la práctica, lo cual incide en beneficios para los alumnos principalmente.

4.3.3.1.5 Impacto

En el inciso anterior se mostró cómo los profesores participantes en el estudio muestran en su discurso evidencia de reflexionar acerca de varios aspectos relacionados con la evaluación del aprendizaje de lengua. Uno de estos es el impacto que tienen los instrumentos o las evaluaciones, en particular los exámenes, en los alumnos.

De esta manera, el código Impacto se definió para describir aquellas citas que impliquen conciencia sobre la posibilidad de un impacto de la evaluación en el alumno.

Impacto positivo y negativo

Dos de las preguntas de la entrevista tenían relación directa con el impacto. Una de ellas era si los profesores pensaban que los exámenes ayudaban o perjudicaban a los alumnos. A esta pregunta Simón responde:

Yo creo que en la mayor medida sí, en la mayoría de los casos los ayuda, porque no nada más se trata o al menos bueno como profesor yo lo veo de esta manera no solo se trata de la calificación, aunque pues para los alumnos es solamente la calificación pero yo creo que a los alumnos les ayuda muchísimo porque tienen algo tangible que les dice cómo están o cómo salieron a lo mejor en estos temas en esta habilidad... (Simón: Entrevista, 7:143-145).

Graciela contrasta con Simón, ella considera que los exámenes contribuyen a que los alumnos no aprendan de manera significativa, sino que se enfoquen únicamente en el examen y en un aprendizaje a corto plazo. Además, enfatiza el estrés que causa en los alumnos.

... pero a veces también siento que es como que un estrés, así como que muy necesario porque se ponen a estudiar tanto, tanto que se les olvida estudiar otra cosa o que nada más se lo aprenden para un momento... (Graciela: Entrevista, 8:130).

Los profesores participantes declaran que los exámenes tienen impacto tanto positivo, como negativo en los alumnos, les ayudan en algunos aspectos, pero también los perjudican en otros, presentando una situación compleja que se debe a varios factores que se presentan a continuación.

Impacto y estrés

Se preguntó a los participantes, de manera directa, cuál era el impacto de los exámenes en los alumnos. Algunos profesores, como Gina (Entrevista, 5:88), Graciela (Entrevista, 8:133) y Belinda (Entrevista, 9:26) comentaron sobre el estrés que provocan los exámenes. Gina, incluso comenta que los exámenes también generan estrés en los profesores:

entonces la verdad que, a mí, el evaluar, pues sí estresa porque pues claro tienes que mostrar todo lo que sabes, o darte cuenta de lo que no sabes, y eso es estresante pues porque no te gusta... (Gina: Entrevista, 5:86).

Una de las respuestas más recurrentes en cuanto al impacto fue el estrés que generan los exámenes en los alumnos, incluso afectando su estado de ánimo, como detalla Graciela. Gina llevó su percepción del impacto a los profesores, que de acuerdo con ella, se ven en la obligación de demostrar conocimiento.

Impacto a nivel personal

Los participantes también hicieron referencia a cómo los exámenes tienen consecuencias en el nivel personal. Erika (Entrevista, 2:98-99) menciona el efecto en los alumnos que tiene un examen a este nivel, cuando se obtienen malos resultados; Nadia (Entrevista, 12:121) considera algunas posibles repercusiones en la vida personal de los alumnos, específicamente a nivel familiar.

Sí, yo creo que pues les generas a veces en la parte como personal de ah no, no sé nada o no aprendí mucho o me hace falta esto o no ese tema de plano no y hay quien de plano no es que ya el inglés no es para mí... (Erika: Entrevista, 2:98-99).

Nadia considera algunas posibles repercusiones en la vida personal de los alumnos, específicamente a nivel familiar:

... entonces sí, yo creo que sí los impactan las evaluaciones y pues obviamente más bien los resultados, porque pues cuando tienen una buena nota a lo mejor los papás hasta los premian y cuando tienen una mala nota los castigan ¿no? En muchos casos. (Nadia: Entrevista, 12:121).

Los participantes del estudio muestran conciencia acerca de la importancia de su trabajo, específicamente de las consecuencias en el nivel personal de sus acciones en el proceso de evaluación y, además de conciencia, hay evidencia de preocupación y empatía.

Impacto, confiabilidad y validez

En algunas ocasiones, los profesores manifestaron como algunas prácticas de diseño de exámenes relacionadas con los aspectos técnicos de estas tienen un efecto. Erika (Entrevista, 2:97) y Nora (Entrevista, 3:180 y 3:204) comentan que la falta de confiabilidad, como reactivos mal elaborados, tiene un impacto en el alumno.

La falta de validez de contenido también tiene repercusiones en los alumnos, como comenta Nora:

...porque sí creo que los resultados y el hecho de que los chicos vean que un tema no lo vimos sí les repercute... (Nora: Entrevista, 3:93).

Nora también apunta que el examen puede tener impacto en la actitud de los alumnos hacia las evaluaciones:

Una desventaja o una manera en lo que los puede afectar es cuando el examen está mal diseñado, cuando un examen no abarca todo lo que tendría que evaluar o abarca más o tiene más, pues sí lo perjudica y creo que lo puede perjudicar a modo que diga, es que odio los exámenes, estuvo horrible, estuvo difícilísimo... (Nora: Entrevista, 3:93).

Se menciona ‘mal diseño’ que pudiese implicar uno o varios de estos aspectos técnicos. Entre los que se implicaron de manera particular y que los profesores manifiestan que pudiesen tener un efecto en los alumnos están la confiabilidad, validez de contenido.

En el análisis se encontró la dualidad negativa o positiva del posible impacto de los exámenes de lengua, especialmente en el alumno. El aspecto positivo sería la contribución al aprendizaje del alumno, sin embargo los aspectos negativos resaltan como el estrés y otros a nivel personal y las consecuencias de las mencionadas fallas en el diseño de los instrumentos. De esta manera se da una relación entre la dimensión ética con la personal y con la técnica.

4.3.3.1.6 Diálogo

En ocasiones durante las entrevistas, los participantes mostraron evidencias de establecer un diálogo con otros integrantes del proceso de evaluación. Para describir esto se creó el código Diálogo, que es cualquier referencia a la comunicación real o hipotética entre el hablante y el alumno. La comunicación real consiste en una narrativa acerca de un diálogo que ocurrió en realidad, mientras que con otras situaciones, el diálogo es hipotético, cuando el participante se expresa acerca de un diálogo general, permanente o consistente. En seguida se muestran ejemplos de diálogo.

Simón comenta que hay un diálogo con sus alumnos al darles a conocer el contenido de los exámenes.

... entonces... sí me gusta incluirlos y me gusta hacerles saber qué va a estar en el examen para que, bueno creo que se los pueda como incentivar para que puedan tener más formas y que no nada más se quede con lo que ya igual han aprendido... (Simón: Entrevista, 7:67).

El diálogo acerca del examen y de los resultados obtenidos, así como dificultad de los reactivos o tareas fueron frecuentes en las entrevistas (Simón: Entrevista, 7:123 y 7:128; Nadia: Entrevista, 12:101-102 y 12:107). En ocasiones el profesor es quien busca retroalimentación del alumno, con respecto al examen, como son los casos de Virginia (Entrevista, 11:32-33), Nora (Entrevista, 3:128) y Camila (Entrevista, 3:128).

La misma Virginia, quien es una de las participantes que evidencian más diálogo con sus alumnos de entre los participantes en el estudio, explica que repite los temas en los exámenes varias veces hasta que todos los alumnos demuestran dominar el tema:

... si una persona no lo domina (el tema) vuelve a aparecer (en el examen) y les es cansado para ellos y me lo dicen es que otra vez, madame pero si ya me lo sé, le digo pues no importa aunque ya te lo sepas tu responde y ya, y sin embargo se vuelven a equivocar, quiere decir que no me lo dominan o que les falta madurez o que les falta práctica y entonces ellos mismos también me dicen, lo voy a practicar más ya lo voy a practicar ¡Ah ahora si ya lo entendí! pero hasta la tercera o cuarta vez. (Virginia: Entrevista, 11:160).

El diálogo entre el profesor y sus alumnos sobre los exámenes se describe en varias vertientes como aspectos técnicos, su contenido, dificultades y errores, resultados, opiniones sobre ellos y aspectos lingüísticos como gramática y vocabulario, y que se relacionan con otras dimensiones.

En esta sección se han descrito los principales componentes de la subdimensión interpersonal, una de las dos de la dimensión ética. La construcción de la interpersonalidad en el discurso de los participantes se da a través de la alteridad, el cuidado, las manifestaciones de exclusión e inclusión, la reflexión, la consideración del impacto y el diálogo. Estas manifestaciones de relaciones interpersonales son parte esencial de la ética del individuo, del profesor en este caso, que, salvo la exclusión, todas implican la construcción o mantenimiento de esas relaciones. En contraste, es preciso señalar la exclusión dentro de este sistema complejo como referente de acciones que generan una separación entre el profesor y los otros participantes debido a la diversidad de pensamiento y de diferencias individuales que se reflejan en la práctica profesional, lo que no contribuye a la construcción de relaciones, pero sí son un indicador a considerar para identificar divergencias que ilustran una dinámica de comportamiento de los profesores.

4.3.3.2 Subdimensión valorativa

4.3.3.2.1 Autoevaluación

En ocasiones los participantes expresaban un juicio sobre con respecto a su propio desempeño y prácticas evaluativas.

Lorena, al explicar que en ocasiones no cubre los temas establecidos en un curso y cómo afecta esto al diseño de exámenes, comenta:

... porque si no yo misma me voy a calificar mal en cierta forma, ¿no? porque no les di todos (los temas)... (Lorena: Entrevista, 1:101).

En ocasiones la profesora se da cuenta de algún aspecto del examen que no funcionó bien por comentarios de los alumnos:

... y sí, cuando veo que todos tienen mal digo ay aquí algo hubo que fue mi culpa, ahí no diseñé bien el instrumento... (Lorena: Entrevista, 1:164).

Lorena (Entrevista, 1:181-182) también evalúa su propio trabajo durante el proceso de diseño de exámenes. En su entrevista, declaró que elaborar un examen le lleva algunos días, lo que le ayuda a revisar y corregir. En ocasiones la autoevaluación llega hasta después de la aplicación de un examen. Como en el caso de Lorena (Entrevista, 1:240-241), quien comenta acerca de cómo se da esta reflexión sobre su instrumento de evaluación:

Otros ejemplos de autoevaluación fueron los siguientes: Erika (Entrevista, 2:55-56) comenta que ella no se siente segura con los reactivos que pudiese elaborar por lo que prefiere tomarlos de libros de texto; Nora (Entrevista, 3:51) explica que evaluar gramática y vocabulario le ayuda a conocer acerca de su enseñanza y su efecto en los alumnos y hace una reflexión sobre su trabajo de docencia y diseño de exámenes en el pasado (Nora: Entrevista, 3:96); De acuerdo con Camila, la experiencia le ha ayudado a elaborar mejores exámenes (Camila: Entrevista, 4:59).

Los profesores se expresan acerca de sus prácticas de diseño de exámenes calificando entre “mal” y “bien” su propio trabajo y en algunos casos manifiestan haber mejorado. Estas reflexiones en que se auto evalúan los participantes son evidencia de una conciencia acerca de su práctica profesional. El docente busca la mejora; realizar su trabajo mejor contribuye a su formación personal y profesional, lo cual redundará en beneficio propio y de los otros involucrados en el proceso evaluativo. Por otro lado, también se encontraron evidencias de juicios sobre éstos, de parte de los profesores, los que no necesariamente contribuyen a la construcción de una ética y responsabilidad profesional:

4.3.3.2 Juicio sobre el alumno

Durante el discurso de los profesores se encontró que emitían juicios sobre la conducta o desempeño de un alumno, lo que se etiquetó con el código Juicio al alumno. Estos juicios se dieron en referencia a varios aspectos relacionados con los estudiantes de lenguas, como su desempeño, habilidades y conducta.

Juicios con base en el desempeño del alumno

Por ejemplo, Lorena hace un juicio del alumno con base en su desempeño en una tarea:

... un alumno puede aprenderse las cosas de memoria, y ponértelas en el examen y sacar diez, pero al día siguiente o a la semana siguiente le pones un proyecto en el que debe utilizar ese conocimiento que ya revisaste, que ya evaluaste y según tú sacó el alumno diez y resulta que no sabe manejarlo... (Lorena: Entrevista, 1:93-95).

Graciela (Entrevista, 8:111) también hace su juicio con respecto a su experiencia en evaluación y observación que ha hecho de los alumnos y para Nora (Entrevista, 3:64-65), la manera en que contestan un examen los alumnos y el tiempo que les lleva hacerlo le sirve para hacer conjeturas:

El que terminen en media hora tampoco es real, quizá no la hora cuarenta minutos de tu clase, sino un examen que refleje cierta complejidad si termina en media hora quizá el chico es muy brillante o no le importa y lo contesto como quiso... (Nora: Entrevista, 3:64-65).

Las profesoras emiten juicios con base en lo que perciben del alumno por sus resultados en un examen, en ocasiones hacen calificaciones con adjetivos, como que son “medio distraídos”, en otras realizan afirmaciones que para ellas parecen explicar el resultado de un examen, como que el alumno “flojea” o que “no le importa”.

Juicio sobre las habilidades de los alumnos

Por otro lado, en algunos fragmentos los profesores hacen conclusiones acerca de las habilidades de los alumnos. Lorena (Entrevista, 1:93-95) emite un juicio sobre las competencias que el alumno supuestamente debería tener, pero, a su decir, no las tienen. En contraste, Graciela (Entrevista, 8:111) comenta que los exámenes son una muestra del logro de objetivos durante el curso

Las citas son una muestra de las conclusiones a las que las profesoras, en este caso Lorena y Graciela, parecen llegar acerca de las habilidades para leer o para entender y realizar algunos ejercicios. Las profesoras parecen expresarse en este sentido con base en sus experiencias.

Juicios con base en las conductas de los alumnos

En cuanto a los juicios realizados con base en las conductas de los alumnos La profesora Virginia se ve a sí misma como una profesora estricta y busca que los alumnos dominen los temas que imparte y concluye su explicación con esta cita:

... y yo se los digo desde el principio si yo veo que no tienes el nivel no vas a pasar (Virginia: Entrevista, 11:185).

Graciela habla en distintas ocasiones sobre justicia en la evaluación y en algunas de esas citas se refiere a los alumnos

... creo que no nada más es que sepan las cosas, sino que sean personas honestas, sobre todo pues por el momento que estamos viviendo, por el país en el que vivimos, entonces me ha tocado que hay alumnos que son algunos alumnos que, pues son así medio, pues que les gusta copiar ¿no? entonces dices: son muy buenos o muy astutos en esa parte, pero en realidad eso no es. Eh, pues no se me hace justo que a lo mejor les vaya muy bien y a otros que lo intentan y pues a lo mejor no lo estudiaron o no lo han aprendido, eh, les vaya mal y los tenga que penalizar por eso, no sé. (Graciela: Entrevista, 8:71-74).

En sus conclusiones sobre la entrevista, Graciela vuelve a hacer referencia a la justicia y agrega lo siguiente:

... porque yo tengo eso muy metido, que a veces la evaluación a veces hay un alumno muy bueno en clase, pero es una persona terrible y pues a veces no, no puedes calificar eso porque ya están formados, ya son en este caso jóvenes adultos... (Graciela: Entrevista, 8:71-74).

La profesora pareciera expresar un deseo de hacer aspectos ajenos a la competencia comunicativa parte de la calificación, sin embargo, ella misma explica que eso no es posible, sobre todo, tratándose de adultos.

Alondra emite un juicio sobre los alumnos que implica, de igual manera, el trabajo de los profesores:

... yo personalmente considero que, al meter todas las habilidades, mi percepción es evaluamos todo, les enseñamos en teoría todo y no dominan nada... (Alondra: Entrevista, 6:12).

Alondra implica que los profesores cumplen con su responsabilidad, pero los alumnos no cumplen la expectativa (de la profesora) que es el dominio del contenido, transfiriendo la responsabilidad de un aprendizaje exitoso al alumno. Las palabras de Alondra emiten un juicio positivo acerca de los profesores, pero negativo acerca de los alumnos.

En la mayoría de las citas en que los profesores realizan juicios acerca de los alumnos puede percibirse una connotación negativa, lo cual contrasta con el discurso del cuidado del alumno que se ha mostrado en otros extractos. Los juicios se emiten, principalmente en razón de las habilidades de los alumnos, su conducta y su desempeño, de los cuales los profesores no parecen tener una percepción favorable.

En este inciso se dio cuenta de los juicios que los profesores emitieron en las entrevistas acerca de los alumnos, sin embargo, se encontraron otro tipo de valoraciones que se describen enseguida.

4.3.3.2.3 Valoración

Durante las entrevistas los profesores hicieron valoraciones de distintos aspectos del proceso de evaluación. A aquellas citas que hacían referencia a esto, se les etiquetó con el código valoración. Dos códigos adicionales se consideraron como componentes a este: Valoración positiva, que son aquellas citas que muestran una valoración de polaridad

positiva acerca de una evaluación; y Valoración negativa, aquellas citas que muestran una valoración de polaridad negativa acerca de una evaluación. En seguida se muestran aquellos aspectos con valoración de parte de los profesores.

Las valoraciones positivas encontradas reflejan dos tipos de opiniones, una con respecto a los exámenes como instrumentos de lengua en general (Gina: Entrevista, 5:27; Camila: Entrevista, 4:108; Belinda: Entrevista, 9:248) y otra en referencia a los exámenes diseñados por los propios profesores participantes (Lorena: Entrevista, 1:141-142; Graciela: Entrevista, 5:27; Genaro: Entrevista, 10:183; Gina: Entrevista, 5:43). En estas últimas, puede observarse modalidad en las expresiones de los profesores con respecto a sus exámenes, que podrían deberse a la interacción con el investigador, durante el momento de la entrevista.

Los profesores no solo hicieron valoraciones o juicios acerca de los alumnos o los exámenes, sino también de sí mismos como docentes. Al contestar la pregunta: “¿Qué nos dicen sus exámenes sobre usted como profesor de lengua?” Gina contestó:

... pues la maestra es buena, la maestra es responsable, la maestra sí es coherente, creo que sí. No sé tendrías a lo mejor que eso preguntárselo más a detalle a mis alumnos. (Gina: Entrevista, 5:93).

Gina califica de manera positiva su labor como docente, tomando como base su trabajo de evaluación, con calificativos como buena, responsable y coherente, terminando su respuesta con un grado de modalidad.

Aunado a lo anterior, también se da una valoración positiva con respecto al diseño de los exámenes, que Graciela (Entrevista, 5:27) califica de “sencillo”. Esto se confirmaría más adelante en la entrevista, cuando contesta a la pregunta sobre qué piensa de sus propios exámenes

Pues ahora que estoy trabajando en la facultad (de Lenguas) me gustan mucho porque son bastante prácticos, son sencillos, son concretos... (Graciela: Entrevista, 8:82).

Graciela reiteró su percepción sobre lo sencillo que es diseñar exámenes, no obstante, fue la única profesora en realizar una valoración en ese sentido, lo cual contrasta con otras opiniones, en las cuáles se expresa incluso duda de parte de los profesores al diseñar un instrumento. Cabe recordar que Graciela fue la participante más joven y con menos

experiencia, lo cual podría explicar su postura, única entre los participantes, hacia los exámenes.

En contraste, el discurso de los profesores muestra una serie de valoraciones negativas. Una de estas opiniones sobre el diseño de exámenes se da en sentido del trabajo que representa diseñar reactivos, al que Graciela califica de “tedioso”, a la calidad del diseño del examen y su impacto en los resultados.

... Ay se me hace muy tedioso (elaborar reactivos), pues es que como que siento que es lo que viene diseñado por lo que vimos o sea exactamente el vocabulario la gramática... (Graciela: Entrevista, 8:94).

Lorena (Entrevista, 1:189) comenta sobre el impacto negativo de exámenes con problemas de diseño y Nora (Entrevista, 3:166) explica de qué manera los alumnos se ven perjudicados con los exámenes

Estas reflexiones de connotación negativa no fueron las únicas que se encontraron en el análisis. Los profesores también se manifestaron en ese sentido, acerca de los exámenes de diseño propio.

Nora explica que impartir la materia de Evaluación le ha ayudado a mejorar sus instrumentos, aun así no tiene una valoración positiva sobre su propio trabajo

... me ha ayudado mucho dar esa materia para mejorar mis exámenes, no por decir que hago unos exámenes confiables, válidos, prácticos, auténticos porque la verdad es que no... (Nora: Entrevista, 3:102).

La profesora manifiesta no diseñar exámenes con varias especificaciones técnicas, sin embargo, en su discurso sí muestra conciencia acerca de ello, así como evidencia de varias prácticas que contribuyen al logro de estas cualidades.

Gina explica que en ocasiones tiene dudas sobre sus instrumentos, por ejemplo, cuando tiene en mente una respuesta a un reactivo y los alumnos contestan diferente. La participante agrega:

Entonces cuando me doy cuenta que no pasa así y me dan otras respuestas y que es la mayoría, si es así de qué hice mal ¿no? ¿En qué fue lo que fallé?... (Gina: Entrevista, 5:37).

A la misma pregunta acerca de las dudas sobre los instrumentos, Simón comenta

... pero no porque realmente se puedan tener esas dos respuestas sino porque fue un error de diseño del examen entonces digo cuando te toca a ti hacerlo dices chin pues también ya la regué ¿no? (Simón: Entrevista, 7:137).

Gina y Simón reconocen haber cometido errores en el diseño de exámenes, lo cual podría significar el inicio de un proceso de responsabilidad y de una conducta ética y profesional.

Algunos profesores hacen una valoración negativa acerca de prácticas evaluativas realizadas previamente, por ellos mismos. Tal es el caso de Nadia, quien confirma un profundo cambio entre su práctica evaluativa al inicio de su carrera docente y el presente:

Ha cambiado obviamente pues el formato, a lo largo de mis años a veces observo de mis bancos anteriores y a veces digo ¿cómo pude hacer eso? (Nadia: Entrevista, 12:66).

De manera similar a Gina y Simón, en las citas previas a ésta, Nadia reconoce y reflexiona acerca de un aspecto en su diseño de exámenes que valora de manera negativa. El reconocer esto podría implicar que en el momento de la entrevista dicha práctica ha cambiado, resultando en un desarrollo profesional.

Otros aspectos del diseño que se valoraron de manera negativa fueron relacionados con las tareas a realizar en un examen:

Virginia comenta que no siempre hace evaluación de la producción oral y explica porqué

... sin embargo, no en todas las evaluaciones hago producción oral porque es muy cansado y eso es negativo para el evaluador... (Graciela: Entrevista, 8:94).

En este caso Virginia se refiere a una práctica que no realiza, debido a que considera que no es positiva para la obtención de un resultado confiable, ya que el desempeño del docente podría verse afectado por la exigencia de la tarea.

Nora relata una ocasión cuando experimentó con un tipo de tarea diferente y los alumnos no tuvieron los resultados esperados:

... la verdad es que sí se los puse (el audio de la tarea a realizar) otra vez y aun así pues no les fue como que tan bien... (Nora: Entrevista, 3:29).

Virginia y Nora hacen valoraciones negativas acerca de algunas tareas, ya que generan dudas en los resultados obtenidos por los alumnos. Esto se relaciona con los aspectos técnicos del diseño de exámenes, en particular la confiabilidad.

También con referencia a la técnica requerida para diseñar instrumentos de evaluación, Graciela se expresa acerca de la preparación que tuvo sobre diseño de exámenes al estudiar la licenciatura

... pues estuvo un poco tediosa porque si vimos bastantes cosas y fue como trabajo en equipo porque éramos bastantes entonces como que llegó un momento en que siento que no nos llegamos a, no trabajamos tan bien en equipo y se me hizo como, no sé muy tedioso muy aburrido y pesado... (Graciela: Entrevista, 8:100-101).

En citas presentadas previamente, se mostró que Graciela no valoraba de manera positiva la práctica de diseñar exámenes. En esta ocasión utiliza la misma palabra “tedioso” para calificar su proceso de preparación técnica.

Al tocarse el mismo tema, preparación técnica en diseño de exámenes, Nadia comenta lo siguiente:

... creo que eso es, sí es importante esa materia (Diseño de exámenes), porque me ha tocado trabajar con colegas que a veces nos mandan exámenes que decimos "¿qué fue lo que hizo?" o "¿Qué está pasando aquí?" ... (Nadia: Entrevista, 12:73).

Nadia expresa sorpresa, en una reacción negativa ante el trabajo de otros colegas. Este disenso podría ser resultado de la preparación técnica que ha recibido la profesora y que le permite emitir un juicio ante el trabajo de otros profesores.

La valoración es un pronunciamiento con implicaciones éticas, que involucra aspectos técnicos, profesionales y personales de los profesores, se da en sentido positivo y negativo. La valoración con polaridad positiva se dio en referencia al trabajo de diseño de los instrumentos, tanto ajenos como propios y en la labor de uno de los docentes. En ese mismo sentido se dio la valoración negativa, aunque además del desarrollo de instrumentos, también alcanza niveles más específicos como el diseño de tareas y la preparación técnica. La valoración negativa supera a la positiva en el número de citas y relaciones con aspectos del proceso de evaluación, aunque esto no necesariamente representa negatividad desde la perspectiva ética, sino que en algunos casos, como cuando los profesores opinan negativamente acerca de sus prácticas en el pasado y afirman que en el presente son diferentes, lo que podría considerarse como un hecho positivo, profesional y ético.

4.3.3.2.4 Obligación

Por Obligación se identificaron aquellas citas que expresan exigencia o deber.

Discursivamente, quedaron marcadas las citas con los verbos “deber” y “tener que”. En este inciso se muestran ejemplos de aquellos aspectos que los profesores entrevistados consideraron como una obligación.

Obligación y aspectos de diseño de exámenes

En cuanto al diseño de instrumentos, por ejemplo, los profesores expresaron obligación de tener claridad en las tareas y en los exámenes (Belinda: Entrevista, 9:150-151 y Erika: Entrevista, 2:32), así como trabajar en el diseño de exámenes de manera colegiada (Alondra: Entrevista, 6:117), ambos aspectos técnicos relacionados particularmente con la confiabilidad de los instrumentos.

Obligación y validez

Genaro declara que se tiene obligación de darle validez a los instrumentos, como puede observar en la siguiente cita:

... entonces tiene que evaluar lo que pues lo que se tiene que evaluar lo que se vio en clase el conocimiento que se le proporcionó al alumno... (Genaro: Entrevista, 10:89).

Genaro, sin mencionar explícitamente el principio de validez, se refiere a ello en su cita con la expresión “evaluar lo que se tiene que evaluar” y hace referencia, implícitamente, a la validez de contenido hablando de “lo que se vio en clase” y que también representa una obligación. Belinda hace una implicación similar al mencionar que es una obligación evaluar las habilidades (Belinda: Entrevista, 9:92) lo que le daría validez a los instrumentos.

Obligación de realizar prácticas deseables sobre diseño de exámenes

En el discurso de los profesores se encontraron evidencias de una obligación que se relaciona con algunas prácticas evaluativas que son recomendadas en la literatura especializada:

Belinda (Belinda: Entrevista, 9:82) relata que en sus estudios de maestría se le hizo la recomendación de pilotear sus instrumentos de evaluación antes de aplicarlos. Nadia

comenta sobre la práctica deseable de evaluar la gramática dentro de un contexto (Nadia: Entrevista, 12:25).

Otro aspecto deseable de la práctica evaluativa que se manifiesta como una obligación es la preparación técnica, como lo expresa la misma Nadia:

... sí creo que es muy importante que en la universidad se propongan este tipo de cursos obligatorios, no facultativos, obligatorios para los profesores, para homologar cómo estamos evaluando a nuestros alumnos, y hacerlo de una manera más como colegiada o académica, que todos estemos en la misma, onda, en el mismo tono... (Nadia: Entrevista, 12:87).

Nora concluye su entrevista haciendo una reflexión sobre la necesidad de prepararse como docente para realizar evaluaciones y concluye con la siguiente cita:

... pero sí, tendríamos que saber cuáles son los aspectos importantes de un examen los que no los podemos pasar por alto... (Nora: Entrevista, 3:215).

Las prácticas deseables que se expresaron como obligatorias por parte de los participantes fueron el pilotaje de instrumentos, la evaluación de la gramática en un contexto y la preparación en cuanto a aspectos técnicos de evaluación. Estas manifestaciones implican una ética de trabajo y desarrollo profesional y técnico por parte de los profesores.

Obligación e institucionalidad

Por otro lado, los profesores manifestaron obligación de realizar algunos aspectos de evaluación, debido a exigencia de su empleador o institución donde laboran.

Nora, expresa en la siguiente cita que evalúa a sus alumnos con exámenes, porque su institución lo pide así

Sí, sí los evaluó con exámenes, bueno una porque es una política es algo que tenemos que hacer entonces está marcado que tenemos que hacer evaluaciones, tres evaluaciones, primer parcial, segundo parcial y la evaluación final. (Nora: Entrevista, 3:1).

Aunado a esta cita de Nora, se encuentra el comentario de Belinda (Belinda: Entrevista, 9:4), donde manifiesta que utiliza exámenes para evaluar porque así se lo piden, se asume que es su empleador quien solicita a la profesora evaluar de esa manera. La respuesta de Belinda no permite distinguir si la obligación de evaluar con exámenes es impuesta o propia, o si es impuesta pero manifiesta categóricamente estar de acuerdo con ello. En el caso de Nora, sí es explícita en cuanto a que es una obligación institucional.

Las expresiones de obligación se relacionan principalmente con aspectos técnicos del diseño de instrumentos de evaluación, con prácticas deseables o recomendables, que se esperan de un profesional del área y contribuyen a la ética del profesor.

4.3.3.2.5 Justicia

El código denominado Justicia se refiere a aquellas citas que expresan conciencia sobre la ausencia de sesgo, favoritismo o perjuicio hacia un individuo en la evaluación.

Inicialmente, se consideraron tres códigos que expresaban ideas similares, justicia, equidad y *fairness*. En este último se utilizó el término en inglés, ya que la idea que se expresa con esa palabra no logra ser reflejada con una sola en español. Finalmente, los tres códigos se fusionaron en el denominado Justicia.

Un aspecto que se implicó en repetidas ocasiones fue la necesidad de buscar justicia para los alumnos a través del logro de la validez de contenido en los instrumentos. Esto en el sentido de evaluar lo que se estudió en la clase y evitar incluir en la evaluación lo que no.

Por ejemplo, Graciela hizo mucho énfasis en ser justos en las evaluaciones. En las siguientes citas, después de explicar ciertas dudas que tuvo sobre algún instrumento, explica:

porque tampoco se me hacía justo que les fuera a calificar algo que ni siquiera habíamos visto bien y no sé cómo que esa parte de la justicia la tengo muy metida... creo que es algo personal, entonces siempre trato de que las cosas sean justas o en medida de lo que a mí me parece justo, entonces si veo que es un ejercicio que pues nadie entendió, que nadie lo tiene bien o la mayoría no lo tienen bien pues es por algo que está faltando y que no puedo ser tan no sé, mala, en ese sentido con los alumnos. (Graciela: Entrevista, 8:66-70).

Para Graciela la justicia parece tener una gran importancia en su práctica, empezando por el contenido del examen. La profesora deja entrever que la justicia puede ser relativa, cuando expresa que trata de que las cosas sean así o al menos como a ella le parece que es justo.

Lorena comenta que sus exámenes han cambiado desde el inicio de su carrera al momento de la entrevista. Entre los aspectos que menciona está lo siguiente:

... creo que las preguntas son más justas con lo que vimos en la clase... (Lorena: Entrevista, 1:53).

Al parecer, Lorena encuentra justicia en que los instrumentos tengan validez de contenido, de ahí la expresión "... más justas con lo que vimos en clase...". En ese mismo sentido se

expresarían Camila (Entrevista, 4:106-107), Gina (Entrevista, 5:61) y Graciela (Entrevista, 8:20)

Simón comenta que en ocasiones las calificaciones de los exámenes no son justas, porque además de la competencia lingüística se califican otros aspectos:

... sí es injusta por ejemplo cuando en la forma de evaluación también se incluye lo que no se trabajó en clase a veces o visitas a auto acceso... (Simón: Entrevista, 7:160).

Varias citas sobre justicia se relacionan con lo que “se vio en el curso” o lo que el alumno “debe saber”. Esto confirma la importancia que tiene entre los profesores la validez de contenido, como se ha visto en otros momentos de las entrevistas presentados con anterioridad.

La validez de contenido es el aspecto que se relaciona más con el de justicia. Sin embargo, hay otros aspectos que se encontraron junto con este último, como poder o abuso de poder (Gina: Entrevista, 5:98), compensación como medio de justicia ante las fallas en el instrumento o errores del profesor (Camila: Entrevista, 4:51), la justicia como medio de igualdad (Nadia: Entrevista, 12:107) y la confiabilidad como un aspecto técnico que asegura la justicia en el examen (Nadia: Entrevista, 12:112).

4.3.3.2.6 Importancia

El Código Importancia se refiere a las menciones explícitas de la importancia de algún aspecto relacionado con la evaluación y diseño de exámenes. En su discurso de los profesores consideraron como importantes los siguientes aspectos.

Importancia de la evaluación

Algunos profesores consideran la evaluación como un aspecto muy importante. En este sentido Nora y Gina mencionan lo siguiente:

... la evaluación y esta cuestión de exámenes es un aspecto tan importante de la docencia que creo que puede ser el momento decisivo en el que puedes perder a tus alumnos... (Nora: Entrevista, 3:104).

... yo creo que tanto la evaluación como el diseño de exámenes son dos temas súper importantes en la práctica y en la formación docente... (Nora: Entrevista, 3:199).

... como conclusión, yo puedo decir que la evaluación es importante, es muy importante, sea utilizando exámenes o utilizando proyectos, ¿no? (Gina: Entrevista, 5:95).

Para Lorena también es importante que se lleve a cabo una evaluación del aprendizaje continuamente.

... entonces, este, creo que sí es importante verificar si de verdad están aprendiendo eso a lo largo del tiempo, porque de un día para otro se pueden acordar, pero después de un mes, de dos meses, ya no es fácil que se acuerden... (Lorena: Entrevista, 1:24).

Al mencionar la relevancia de la evaluación en la práctica docente, los profesores agregan algunos conceptos como el diseño de exámenes, proyectos y el aprendizaje de los alumnos.

Importancia de la preparación técnica en evaluación

Se les preguntó a los profesores si consideraban importante la preparación técnica en cuanto a diseño de exámenes. Algunas profesoras como Nora (Entrevista, 3:114), Camila (Entrevista, 4:109) y Nadia (Entrevista, 12:80) comentaron que es importante prepararse en cuanto a aspectos técnicos de la evaluación. Podría pensarse que la idea fue inducida por el investigador, pero si bien estas profesoras corroboran la idea de la pregunta, Alondra (Entrevista, 6:58) opina lo contrario expresando que no lo es, pero que sí debería serlo.

Importancia de la validez del instrumento

Definir lo que se va a evaluar, que repercute directamente en la validez del instrumento, se mencionó en las entrevistas. Camila, por ejemplo, concluye su entrevista con la siguiente cita:

... o también cual es el propósito del examen también, si es un tema bastante importante y bastante amplio... (Camila: Entrevista, 4:118).

Lorena (Entrevista, 1:2 y 1:5) y Nora (Entrevista, 3:4) consideran importante medir el aprendizaje con instrumentos diferentes a los exámenes, pero sobre todo que sean válidos. Para Simón es importante que el instrumento tenga validez de contenido (Simón: Entrevista, 7:11).

La validez se confirma en el discurso de los profesores como un aspecto técnico importante de la evaluación. Como en otros apartados del análisis, es notable que los profesores se refieren a este principio, aunque no precisamente por su denominación (con excepción de Lorena), sino que se refieren a la importancia de tener claridad en el propósito del examen o de cumplir con el contenido de los programas.

Importancia del trabajo colaborativo en la elaboración de instrumentos

Erika (Entrevista, 2:112) considera que el trabajo colaborativo es muy importante para mejorar el diseño de exámenes. En otros momentos de la entrevista Erika mencionó cómo se involucra su coordinadora y la manera en que ésta organiza el trabajo entre las profesoras. Erika hace énfasis en la importancia de este trabajo y el papel de la coordinadora. Este tipo de reflexión fue única entre los participantes.

Importancia de la gramática y el vocabulario

Nora (Entrevista, 3:46-47 y 3:52) y Simón (Entrevista, 7:64) comparten su creencia en la importancia del vocabulario. Ella también habla de la importancia de la gramática y en su caso, no solamente de la manera en que usen este aspecto de la lengua los alumnos, sino como parte de una reflexión sobre su propio desempeño.

Importancia de la formación personal

Graciela menciona que para ella el desarrollo personal de los estudiantes es importante:

Creo que también tienen que formarse como personas, entonces creo que es importante no sé, que sean responsables de lo que están haciendo durante el semestre o el curso para que pues aparte de de aprender conocimiento tengan también crecimiento como personas... (Graciela: Entrevista, 8:9-11).

Para la profesora, no solamente son importante la evaluación o el aprendizaje de la lengua, sino el desarrollo, formación o crecimiento personal, lo cual también involucra un cuidado del alumno.

Los profesores participantes han hecho mención explícita de varios aspectos relacionados con la evaluación de lenguas, como la evaluación en general, la validez, la preparación técnica, el trabajo colaborativo, la gramática, el vocabulario y la formación personal. Las referencias explícitas a estos aspectos también son un reflejo de la responsabilidad de los profesores, de su cuidado a los alumnos, de atención a su desempeño y ponderación de la relevancia de ciertos aspectos de la evaluación, lo que en conjunto contribuye a los aspectos éticos del ejercicio profesional.

La dimensión ética de la responsabilidad profesional puede ser comprendida desde dos subdimensiones, que reflejan los aspectos éticos en el discurso de los profesores de lengua

acerca de la evaluación. Estas subdimensiones son la interpersonal y la valorativa. La primera involucra aquellos aspectos relacionados en cómo se dan las relaciones interpersonales en el proceso evaluativo y cómo el profesor las aborda, cuales son las percepciones que derivan de ellas y el efecto en su ética y su práctica profesional. En el discurso pueden apreciarse como códigos que describen estas relaciones interpersonales la alteridad, el cuidado del alumno, los principios de inclusión y exclusión, la reflexión, el impacto y el diálogo. La segunda subdimensión, la valorativa, es resultado de una apreciación del valor de distintos actores y factores del proceso, que realiza el profesor y hacen resaltar algunos de esos aspectos o actos y acciones de parte de los participantes en el proceso evaluativo. Los aspectos que se consideraron como valoraciones que se encontraron se refieren a: la autoevaluación que hace el profesor, juicios sobre alumnos, valoraciones, obligaciones, justicia e importancia.

A lo largo de este análisis se han presentado las relaciones o conexiones entre esta dimensión y otras identificadas como componentes de la responsabilidad profesional. Se puede percibir una estrecha relación, especialmente, con los aspectos técnicos de la evaluación, es decir, la dimensión técnica. La mayoría de los aspectos técnicos repercuten directamente en la validez de los instrumentos y se encontraron conectados con los códigos de la dimensión ética. Esto nos da un indicio de cómo la ética profesional del profesor se ve reflejada en su práctica, creencias o conciencia, todas evidenciadas en el discurso.

Desde la perspectiva de la complejidad pueden verse dos dimensiones entrelazadas, por medio de diversas conexiones entre sus códigos. Es posible observar cómo las prácticas técnicas de evaluación se ven influenciadas por conductas de carácter ético, que buscan el desarrollo profesional y bienestar del individuo, que conforman lo que se ha denominado responsabilidad profesional.

4.3.4 La dimensión personal

Esta dimensión se utilizó para señalar elementos discursivos que permiten mostrar aspectos personales que difieren de los técnicos, docentes, éticos y que además permiten dar cuenta de la modalidad o cómo se relaciona el hablante, de manera personal, con el enunciado.

La dimensión personal incluye códigos como Opinión (“yo creo”), Preferencia (“me gusta”) e Intención (“trato”, “intento”), así como otros códigos de menor presencia como Imposibilidad o Inconformidad, y otros, que harían un total de veinte.

4.3.4.1 Opinión

Se consideró como una opinión todas aquellas citas que incluyeron el marcador discursivo “yo creo” o “creo”. Este código se interrelaciona con otros cinco: Alteridad y Cuidado del alumno, de la dimensión ética; y Validez, Confiabilidad y Tareas, de la dimensión técnica. Los marcadores de discurso mencionados para este código atenúan la fuerza del enunciado, dando posibilidad al interlocutor de cuestionar la aseveración, o de emitir una opinión distinta. La importancia de esta dimensión radica en

En seguida se muestran algunos ejemplos de la presencia de este código en el discurso de los profesores.

Opinión y Validez

Lorena comenta que sus exámenes han cambiado del inicio de su carrera al momento de la entrevista, y agrega que son más justos, porque incluyen los contenidos de la clase:

... creo que las preguntas son más justas con lo que vimos en la clase... (Lorena: Entrevista, 1:53).

Nora también da una opinión acerca de lo que se debe evaluar

... yo creo que las características de un examen debe (*sic*) de ser del nivel de lengua, debe ser congruente con el nivel de lengua... (Nora: Entrevista, 3:53).

Opinión y Confiabilidad

Camila opina que entre las características que deben tener los exámenes se encuentra algunos aspectos de confiabilidad:

Creo que el examen tiene que ser muy específico en cuanto a los temas, a las instrucciones, al puntaje que se le tiene que dar a cada reactivo... (Camila: Entrevista, 4:38).

Alondra opina que un tipo de tarea, como la opción múltiple, no permite obtener resultados confiables de algunas habilidades

...entonces creo yo que es un poco difícil medir el avance en las habilidades solo por opción múltiple... (Alondra: Entrevista, 4:38).

En la siguiente cita, Graciela opina sobre lo poco justo que es para los alumnos que el profesor tenga errores en el diseño de un examen. Esta cita también contiene los códigos Tareas, Inclusión y Alteridad.

...y yo creo que no sería justo que por algún ejercicio que a lo mejor no está bien diseñado o que no hayamos explicado bien o no esté bien entendido entonces les vaya mal... (Graciela: Entrevista, 8:81).

Opinión y cuidado del alumno

En cuanto al cuidado del alumno, en algunas ocasiones los profesores participantes se expresaron emitiendo una opinión. Tal es el caso de Graciela, quien comenta que es necesario evaluar a los alumnos con instrumentos alternativos a los exámenes y argumenta la siguiente razón:

... creo que también tienen que formarse como personas, entonces creo que es importante no sé, que sean responsables de lo que están haciendo durante el semestre o el curso para que pues aparte de, de aprender conocimiento tengan también crecimiento como personas... (Graciela: Entrevista, 8:9-10).

Probablemente, Graciela explica lo anterior bajo el supuesto que la evaluación alternativa ofrece la posibilidad de evaluar otros aspectos más personales de los alumnos, ajenos a lo que ella llama “conocimiento”.

Opinión y alteridad

En ocasiones los profesores expresaban su opinión acerca de ‘otros’, en particular sus alumnos, y es esta suposición acerca del ‘otro’ la que conjuga la opinión con la alteridad, como se ve en el ejemplo siguiente:

Nadia se refiere a los alumnos, en particular a la estrategia de parafraseo, que ayuda a los alumnos a expresar sus ideas, y a la pregunta del entrevistador sobre si esta estrategia es válida para un examen, la profesora contesta lo siguiente:

Sí, yo creo que sí, porque eso denota una habilidad en ellos. Denota que saben describir, que saben explicar y eso los va a llevar finalmente a comunicarse con las personas... (Nadia: Entrevista, 12:23).

... sí, yo siento que el hecho de asociar pues el contexto de los temas a la gramática sí los puede beneficiar, sí, en eso sí yo creo... (Nadia: Entrevista, 12:33).

La importancia de esta dimensión radica en que la fuerza de la enunciación mengua por la modalidad, probablemente motivado por dudas en el profesor, o una idea de que su declaración es de índole personal, pudiese carecer de un sustento sólido. Como se ha visto en el análisis del discurso, algunos profesores tienen poca experiencia docente o poca preparación sobre aspectos técnicos de evaluación y resulta en que el entrevistado cuide el tono o fuerza de sus pronunciamientos.

Cabe la posibilidad que la modalidad también sea causa del interlocutor, en este caso, el investigador. Es decir, la figura de éste como especialista del área de estudio pudiese causar que las afirmaciones de los profesores fuesen enunciadas con moderación y cuidar su imagen ante este entrevistador.

La modalidad en el discurso contribuye en la práctica educativa a nivel personal, donde el participante modula su enunciado por una situación personal, contextual, posiblemente relacionada con sus antecedentes o con la preservación de su imagen ante el interlocutor. El impacto en la evidencia de la responsabilidad profesional en el discurso no es de las proporciones de las dimensiones ética, técnica o lingüística, pero se consideró dar lugar a esta dimensión para dar cuenta de este efecto personal en la construcción de la responsabilidad profesional.

4.3.5 La dimensión docente

Esta dimensión consiste en 16 códigos que describen prácticas docentes concurrentes al diseño de exámenes, como otros tipos de evaluación, léase evaluación continua, evaluación alternativa; otras prácticas como el uso del libro de texto y algunas políticas institucionales que rigen la práctica docente de los profesores. En este inciso se presentan estas prácticas agrupadas en sus respectivos códigos, entre los que resaltan Evaluación Alternativa, Evaluación continua, Práctica docente, Libro de texto y Política institucional.

4.3.5.1 Evaluación alternativa y evaluación continua

Varios de los participantes se manifestaron a favor del uso de instrumentos alternativos, u otros tipos de evaluación del aprendizaje de lenguas, distintos a los exámenes y las calificaron de necesarias (Erika: Entrevista, 2:114), sin embargo, solamente Nora (Entrevista, 3:4-5).y Virginia (Entrevista, 11:2) manifestaron haber puesto otro tipo de

evaluaciones en práctica, a lo que esta última calificó de atrevimiento, cuando narra una ocasión en que evaluó con un portafolio de evidencias. Los exámenes son calificados como poco objetivos, poco válidos o tradicionalistas (Graciela: Entrevista, 8:159).por algunos de los participantes, pero es claro que es la práctica predominante.

Lorena expresa que la evaluación alternativa podría contribuir a dar más validez a la evaluación:

... es muy complicado y es muy difícil que (un examen) llegue a ser objetivo y que mida lo que de verdad sabes, creo que es muy muy complicado necesitaríamos, no sé, no solamente un examen, sino mucho otros elementos para identificar, para poder medir con más certeza... (Lorena: Entrevista, 1:259-260).

Por otro lado, algunos profesores manifestaron evaluar de manera continua durante el curso:

... durante todo el curso en ciertas actividades muy sencillas que les pongo me doy cuenta del desempeño de los alumnos... (Camila: Entrevista, 4:53).

Virginia declaró realizar un “control continuo... por medio de *quizzes*” a lo largo del curso, aunque no define con claridad la diferencia entre éstos y los exámenes:

... la comprensión oral, la producción oral, la producción escrita y comprensión escrita en control continuo también, o sea no nada más llevo exámenes, sino un control continuo por medio de quiz... (Genaro: Entrevista, 10:140).

Varios profesores se expresan acerca de la evaluación continua, sin embargo, en algunos casos parece ser que la denominada evaluación continua son solamente exámenes o *quizzes*, como les llaman algunos profesores, que se realizan con más frecuencia y que son de menor extensión que los exámenes correspondientes a una evaluación sumativa.

4.3.5.2 Libro de texto

Los libros de texto también son utilizados por los profesores para realizar exámenes, ya sea como base para los reactivos, el contenido, o para utilizar los exámenes diseñados por los autores del libro. En seguida se muestran las menciones este en el discurso de los profesores.

Lorena (Entrevista, 1:15) comenta que para la elaboración de reactivos combina algunos ya existentes que toma de los libros de texto con otros de diseño original. En una de las instituciones donde labora, Erika se vale de otras estrategias para el diseño de reactivos

Los libros que teníamos ahí (en Facultad de Ingeniería) no tenía... para hacer exámenes, uno tenía que buscar, por ejemplo, en otras bibliografías, o en internet, en algunas otras fuentes... (Erika: Entrevista, 2:29).

Virginia hace uso de los materiales auditivos de los libros de texto, para diseñar exámenes:

Los audios los tomo del de manuales que ya existen que no estamos llevando en clase pero que tiene afinidad con la temática... (Virginia: Entrevista, 11-40).

Parece ser que Virginia se refiere al libro del profesor correspondiente al libro de texto con la palabra “manuales”.

Independientemente de las denominaciones que dan los participantes al libro de texto, la evidencia es clara acerca de su importancia en la práctica evaluativa.

Por otra parte, el libro de texto tiene un impacto adicional, también suele ser la base del contenido de los exámenes, de esta manera el código Libro de texto se relaciona con la validez de contenido. Como en la siguiente cita donde Nora describe como diseña sus exámenes, proceso que inicia de la siguiente manera.

... entonces, siempre veo el libro y digo: bueno, pues abarqué de la unidad 1 a la 4, a ver qué temas gramaticales son... (Nora: Entrevista, 3:15).

Al definir qué es validez de contenido, Alondra explica:

Los contenidos que se apeguen ¿no? Que estemos evaluando realmente lo que trabajamos, de acuerdo al contenido del programa y los libros de texto ¿no? ehm eso es lo del *content*... (Alondra: Entrevista, 6:164).

Genaro confirma que el contenido del examen está basado en el libro de texto, pero su percepción es que es una limitante para el aprendizaje de los alumnos:

Las desventajas es que nos enfrascamos en la primera (sic) en el libro, es el vocabulario del libro y si bien es completo porque el libro estaba muy completo nos estamos limitando a evaluar solamente ese vocabulario... (Genaro: Entrevista, 10:35).

Como puede observarse, en varios de los casos, el libro de texto juega un papel relevante en la práctica evaluativa, ya sea como fuente de reactivos o tareas, o como base para el contenido de la prueba. Algunos profesores como Nadia y Virginia hacen uso de libros distintos a los que utilizan en su curso y otros docentes, como Lorena, adaptan contenidos del libro para sus instrumentos.

4.3.5.3 Política institucional

La práctica docente y la evaluativa se ven impactadas por políticas establecidas por las instituciones, las administraciones y las jefaturas o coordinaciones, en las que los profesores participantes se desempeñan.

El uso de exámenes como instrumento de evaluación es, a decir de algunos profesores, resultado de una imposición o exigencia de parte de las instituciones donde laboran,

Sí, lo hacemos (evaluar con exámenes) pues respetando lo que es la legislación, lo que se propone en la Universidad, que es con evaluaciones parciales... (Nadia: Entrevista, 12:1).

Sí, si los evaluó con exámenes, bueno, una porque es una política, es algo que tenemos que hacer entonces está marcado, que tenemos que hacer evaluaciones, tres evaluaciones, primer parcial, segundo parcial y la evaluación final. (Nora: Entrevista, 3:1).

En estos casos parece ser que los profesores optarían por una alternativa a los exámenes, ya que todos especificaron utilizar estos instrumentos por obligación institucional.

En algunos casos, el tipo de tareas también resulta ser una imposición. Alondra explica que en la Facultad de Medicina se aplican únicamente exámenes de opción múltiple y explica la razón:

Porque así nos lo nos lo piden. En todas las materias son de opción múltiple, porque obedece a que ellos manejan un sistema de lector óptico de evaluación, o sea nosotros como maestros no calificamos, existe un departamento de evaluación como tal, allá... (Alondra: Entrevista, 6:41).

El tiempo que se les da a los profesores para reportar los resultados de los exámenes también tiene impacto en su práctica:

... entonces, pues es mucho trabajo y para hacerlo más rápido y también para la parte de calificar, porque también te dan un tiempo límite para calificar y subir calificaciones, en este caso... (Graciela: Entrevista, 8:106).

Erika explica que los exámenes se realizan con base en lo especificado por la coordinadora de la Facultad donde trabaja:

... entonces, se iban adaptando (los exámenes) pero lo que a... la coordinadora le gustaba era que llevaran como una lógica, como que los hiciéramos así, como decirlo o sea que llevaran el mismo tema, que no estuviera saltando de una cosa a otra... (Erika: Entrevista, 2:30).

Lorena comenta como el trabajo de diseño de exámenes es organizado por un órgano colegiado en la Facultad donde trabaja:

... en la Facultad de Derecho, nada más uno (examen) de todos, porque ahí la academia hace que todos los maestros hagan varios exámenes... (Lorena: Entrevista, 1:31).

En esta sección se da cuenta de cómo algunas de las prácticas evaluativas son decididas a nivel directivo o institucional, como el tipo de instrumentos o tareas a aplicar. Así como otras políticas que impactan en las decisiones de los profesores como el tiempo para entregar calificaciones o el proceso de diseño.

La dimensión docente conjuga factores de diversa índole como tipos de evaluación, factores internos y externos relacionados con materiales y con tomas de decisiones, que se relacionan con el quehacer docente de manera concurrente a la evaluación. A diferencia de las dimensiones lingüística, técnica y ética, la toma de decisiones acerca de los procesos evaluativos tiene influencia de otros factores descritos en esta sección. No obstante, se ha considerado como parte de la responsabilidad profesional, por la naturaleza de dichos factores, por ejemplo, el profesor obedece a lo señalado por sus superiores o a las políticas establecidas a nivel institucional, lo que bien podría considerarse como práctica profesional, deseable, generando de esta manera una relación con la dimensión ética.

También se encontraron conexiones con otras dimensiones. La dimensión lingüística se relaciona con esta por medio de las consideraciones a los contenidos gramaticales, temáticos y léxicos del libro de texto. La relación con la dimensión técnica pudo observarse con relación a los tipos de tareas, aspecto relacionado con la confiabilidad, la validez de los contenidos y la validez de los instrumentos expresada en diversos momentos de manera explícita e implícita en las entrevistas.

De esta manera se han presentado las cinco dimensiones encontradas en el análisis del discurso de los profesores participantes. Se considera que tres de estas son las que tienen un mayor peso en el concepto de responsabilidad profesional, que son la técnica, ética y lingüística, y en menor grado las dimensiones personal y docente.

Dado el gran número de códigos identificados en el análisis, en las secciones anteriores se presentaron solamente los de mayor relevancia, identificados por su presencia y por las relaciones con otros códigos o dimensiones.

El reporte de resultados se presentó conforme a los códigos, categorías y dimensiones que se definieron durante el análisis. Enseguida se presenta el análisis con relación a las preguntas realizadas en las entrevistas.

4.4 Análisis del discurso por temas

En esta sección se dará cuenta de los resultados obtenidos acerca de las prácticas evaluativas descritas por los participantes en la encuesta. Se consideraron las preguntas realizadas por el investigador como puntos de enfoque para estructurar para análisis, de manera que cada pregunta representa un tema. El objetivo de este apartado del trabajo es identificar prácticas comunes, coincidencias, así como excepciones y variantes a las mismas.

4.4.1 Proceso de diseño de exámenes

Se les pidió a los profesores describir, paso a paso, el proceso que seguían para elaborar exámenes. Las respuestas fueron en su gran mayoría extensas, pero podrían resumirse en las siguientes prácticas.

Nueve de los doce profesores manifestaron iniciar con una revisión de contenidos. Entre paréntesis se muestra el número de la cita (Lorena, 1:6; Erika, 2:25; Nora, 3:14; Alondra, 6:71), temas (Camila 4:17-19; Simón 7:25; Virginia, 11:39) o de “lo que se vio” o “hemos visto en clase” (Graciela 8:14; Gina 5:4).

Las habilidades comunicativas fueron mencionadas constantemente en respuestas a distintas preguntas durante la entrevista, en esta pregunta en particular siete profesores hicieron mención como objeto de evaluación en el proceso de diseño de exámenes de las “habilidades” en general (Erick, 2:28; Simón, 7:27), agrupadas en receptoras y productivas (Lorena, 1:10; Genaro 10:47), pero principalmente de todas o alguna de ellas en particular (Lorena, 1:7; Nora, 3:19; Camila, 4:18; Alondra, 6:79; Simón 7:28-29; Genaro, 10:49; Virginia 11:38). Las habilidades son mencionadas con frecuencia, junto con el tipo de tareas que se aplican para ponerlas a prueba (Simón, 7:27; Genaro, 10:65). Algunos profesores mencionaron que estructuran sus exámenes con base en las habilidades a evaluar

(Simón, 7:27; Belinda, 9:60; Virginia, 11:38). Otro tema recurrente entre los profesores fue la necesidad de buscar un balance entre las habilidades a evaluar (Ericka, 2:28; Alondra, 6:79; Lorena, 1:7 y 1:10).

Los profesores expresaron que en ocasiones utilizan el libro de texto para el diseño de los exámenes. Algunos adaptan contenidos o tareas de su libro de texto o de otros del mismo nivel (Camila, 4:21 y 127; Nadia 12:37; Virginia 11:40). Otros profesores utilizan los exámenes diseñados por los editores y que suelen acompañar al texto (Graciela 8:12; Nadia, 12:34) o como Ericka (2:29) quien manifestó que lamentaba que el libro de texto no los incluyera. Por otro lado, Nora (3:17) hizo énfasis en que no utiliza este tipo de exámenes, al menos los que evalúan vocabulario y gramática.

Al describir su proceso de diseño de exámenes, los profesores mencionaron que utilizan cierto tipo de tareas como Alondra (6:84) que menciona preguntas abiertas, opción múltiple, *fill in the blanks* y *cloze*; Genaro menciona *Fill in the gap* (10:56) y falso-verdadero (10:51).

Algunos profesores expresaron que buscan que las tareas a realizar por los alumnos sean familiares para ellos (Alondra, 6:74; Simón, 7:34; Belinda, 9:63).

En resumen, el proceso de diseño de exámenes inicia con una revisión de los contenidos del curso, para después determinar las habilidades a evaluar y diseñar o adaptar tareas con base en el libro de texto u otros materiales similares, con algunas variaciones en cuanto a los tipos de tareas. Por supuesto, cada proceso varía de acuerdo al individuo y el centro de trabajo donde se desempeña el profesor y otras variables mencionadas a lo largo de este documento, como la experiencia, formación, capacitación, desarrollo profesional y creencias.

Esta diversidad de factores, que son identificados en códigos y dimensiones en este trabajo, y cómo influyen en las decisiones que toma el profesor durante el proceso, ejemplifican cómo se comporta un sistema complejo, donde cada profesor es un individuo único, irrepetible y que estas características lo llevan a tomar decisiones que contribuyen al logro del diseño de un instrumento que es a su vez parte de un sistema mayor. Estos individuos se integran y sus decisiones y prácticas contribuyen al logro general, se adaptan a su contexto,

incluso como pudo verse en el caso de varios de los participantes, se desempeñan en distintas instituciones dentro de la misma universidad y tienen prácticas diferentes. No obstante, el profesor en esa adaptabilidad, conjuga esa variedad de factores en el logro del diseño del instrumento.

4.4.2 Características de los exámenes de lengua

Se les preguntó a los profesores cuáles son las características con que debe contar un examen de lengua. A pesar de que durante la entrevista se mencionaron varias, en este análisis se presentan las que fueron expresadas por los participantes a solicitud explícita del entrevistador.

Los profesores mencionaron como características la confiabilidad (Camila, 4:44; Gina, 5:24) y la validez (Genaro, 10:76), utilizando los términos técnicos, aunque en un par de casos (Camila; Genaro) no se utilizaron de acuerdo con las definiciones establecidas en la teoría en la materia.

Otra característica recurrente fue el equilibrio o balance de las tareas y del grado de dificultad en un examen. Los profesores se refieren a que debe existir un balance entre las habilidades a evaluar (Lorena, 1:41), incluyendo la gramática. Gina (5:30) mencionó la confiabilidad, sin embargo, explica que con el uso de ese término se refiere a que el examen sea apto al nivel de sus alumnos, lo que sería en realidad, validez. Aida (6:162) se refirió a los términos técnicos en inglés, por ejemplo, *face validity* o validez visual, *content validity* o validez de contenido y *reliability* o confiabilidad. En el mismo sentido, otros profesores (Nora, 3:53; Nadia, 12:45) expresaron que los exámenes deben ser congruentes o correspondan al nivel de lengua del alumno, lo cual también podría considerarse como validez.

Otras características mencionadas fueron que en el examen se evalúen todas las habilidades (Nora, 3:54; Camila, 4:39; Gina, 5:31; Belinda, 9:92); que en el examen se evalúe el contenido o temas que se hayan visto o trabajado (Lorena, 1:36; Gina, 5:26; Simón, 7:80); que el examen represente un reto para el alumno como expresan Nora (3:59) “deben de tener cierta complejidad”, Ericka (2:68) y Graciela (8:49), quien utiliza la palabra inglesa ‘*challenging*’.

Los profesores también mencionaron como características de los exámenes de lengua, que el examen tenga validez visual o *face validity* (en inglés). Aida (6:162) es la única que utiliza el término técnico, en este caso en inglés; ideas similares son expresadas por Genaro (10:76), “tiene que parecer un examen”, y por Ericka (2:66): “debe ser amigable con un formato que no sea pesado”, o Lorena (1:42) quien menciona que el examen en su “diseño visual” debe ser ordenado.

Los profesores expresaron que el examen debe incluir tareas que estén contextualizadas, relacionadas con los temas estudiados en el curso (Nora, 3:55; Nadia, 12:45); así como incluir una variedad de tareas (Belinda, 9:93; Virginia, 11:113).

El examen debe, además, ser claro (Ericka, 2:66; Nadia, 12:45; Virginia 11:112); con instrucciones claras y precisas (Lorena, 1:43; Camila 4:38; y Belinda, 9:94); y ser fácil o práctico para calificarse (Lorena, 1:39; Graciela, 8:52; y Genaro, 10:79).

De esta manera, se concluye que las características que deben tener los exámenes de lengua son: Confiabilidad, validez, equilibrio o balance, evaluar todas las habilidades, validez visual, validez de contenido, debe ser amigable, ordenado, retador, claro, práctico, fácil de calificar, y con instrucciones precisas. Durante la entrevista hubo más implicaciones, sin embargo, como respuestas directas a la pregunta realizada explícitamente por el entrevistador, esas fueron las características mencionadas.

Las características que surgen del análisis del discurso no son solamente aspectos técnicos de diseño de exámenes, sino que involucra la dimensión ética, especialmente aquellas cualidades relacionadas con el ‘otro’, el alumno, en este caso. Varias de las características mencionadas como la validez visual, o el que el examen sea ‘amigable’, ‘retador’ o ‘claro’ implican el bienestar del alumno, un rasgo de carácter ético.

4.4.3 Los exámenes como instrumentos de evaluación del aprendizaje de lenguas

En la entrevista se les preguntó a los profesores qué pensaban de evaluar a los alumnos con exámenes, lo cual permitió conocer algunas percepciones e ideas.

Cuatro profesores fueron explícitos al manifestar su gusto por los exámenes, como Erika, (2:2), Virginia (11:6) y Simón (7:4). Cabe mencionar que Erika, en otro momento de la

entrevista manifestó lo contrario, es decir, que no le gustaba hacer exámenes. Otros como Genaro (10:31), quien expresó admiración por los exámenes de Cambridge Assessment, especialmente para evaluar gramática. En contraste, algunos enfatizaron la importancia, necesidad y utilidad de los exámenes (Nadia, 12:4; Nora, 3:4; y Lorena, 1:1), aunque hicieron énfasis de que no es la única forma de evaluar, puntualizando la evaluación alternativa o continua. Incluso Graciela (8:3) manifestó sentir pereza por los exámenes, debido a lo que representa calificarlos.

Cinco de los profesores expresaron dudar sobre las cualidades de los exámenes de ser válidos y confiables, para lo cual usaron expresiones como “... no es objetivo o cien por ciento objetivo para mí...” (Gina, 5:2), “... de repente miden lo que tienen que medir...” (Graciela, 8:3) o “... la calificación que ellos obtienen no refleja totalmente su conocimiento...” (Erika, 2:2). En el caso de Nadia, comenta que el estrés afecta la calificación de los alumnos y por lo tanto no es tan objetiva (Nadia, 12:8).

Por otro lado, Simón comenta que “... a veces siempre esta esa idea de decir que... a veces el examen no refleja lo que el alumno sabe” (Simón, 7:5), pero contrasta explicando que se debe tener cuidado con el diseño para que eso no suceda y declara estar en favor de los exámenes como instrumento de evaluación. Además, Simón es el único profesor que señala que los exámenes son para medir el aprendizaje. Algunos otros profesores dieron muestra de considerar los exámenes como instrumentos válidos o razones que harían pensar en los exámenes como tales. Por ejemplo, Nadia (12:4), Simón (7:6), Genaro (10:28), Virginia (11:6) y Alondra (6:11) expresan que los exámenes sirven para medir conocimiento y habilidades. Camila (4:10) comenta que ayudan a “puntualizar las diferentes habilidades y temas”, en el sentido que el contenido de los cursos es extenso y los exámenes ayudan a delimitarlo.

Entre las respuestas obtenidas, se expresa gusto, admiración y necesidad por los exámenes, Por otro lado, también se encontraron expresiones no tan positivas como tedio, pereza, duda y falta de objetividad, validez y confiabilidad. Sobre los objetos de evaluación, se mencionaron las habilidades, el aprendizaje y conocimiento.

Las opiniones sobre los exámenes como instrumentos de evaluación se encuentran en un amplio espectro, desde muy positivas, hasta muy negativas, incluso expresadas por un mismo individuo, aunque en cuanto a diferentes aspectos del examen.

La decisión de utilizar exámenes como instrumentos de medición del aprendizaje es institucional, como puede notarse en varios de los contextos ilustrados. Algunos profesores manifestaron esto y precisando a la vez que es posible evaluar de maneras diferentes. Nuevamente nos enfrentamos a una situación compleja y que en este caso involucra grados de literacidad en evaluación de los participantes en el proceso, quienes toman decisiones a nivel institucional, léase facultad, coordinación o academia, y a los profesores. Parece haber disenso en varios de los casos presentados.

4.4.4 Objetivo de los exámenes

Se preguntó a los profesores qué información se podía obtener sobre sus alumnos con base en sus exámenes. La razón de esta pregunta fue para indagar acerca de los objetivos que tiene un profesor para su examen, lo que a su vez nos daría un indicio sobre su validez.

Los profesores declararon que la información acerca de sus alumnos que proporcionan sus exámenes son el nivel de lengua (Camila, 4:84; Nora, 3:138), si se han cumplido los objetivos (Ericka, 2:106; Alondra, 6:200), conocimiento (Lorena, 1:106; Belinda, 9:166; Simón, 7:107), comprensión (Ericka, 2:105; Nadia, 12:88) y dominio (Alondra, 6:209; Virginia, 11:152). Cabe mencionar, que tres profesores (Belinda, 9:166; Alondra, 6:205; Nora, 3:139) expresaron que existen incongruencias entre las calificaciones obtenidas en los exámenes y el desempeño habitual de los alumnos.

Otras respuestas, que no fueron mencionadas por más de un profesor, fueron las siguientes: Lorena (1:104) mencionó que se mide competencia, Simón (7:111) explicó que el examen muestra las estrategias que utilizan desarrollan los alumnos para la resolución de las tareas; Graciela explica que el examen refleja la personalidad de los alumnos y agrega “No nada más se trata de lo que saben sino de lo que son” (Graciela, 8:118). Belinda (9:166) menciona “avance” y Nadia (12:92) explica que el examen muestra aspectos personales como interés o motivación. Nora (3:144) menciona que el examen muestra áreas de oportunidad y fortalezas y Simón (7:110) que el examen ayuda al alumno a hacer una

reflexión sobre su aprendizaje. Lorena (1:99) comenta que si el examen no está bien diseñado “no te deja saber gran cosa”.

Al contestar esta pregunta, cinco profesores emitieron juicios generales sobre los alumnos tanto positivos como negativos, relacionados con su desempeño (Nora, 3:142-143) respuestas y atención (Ericka, 2:104) o aspectos particulares como actitud (Simón, 7:108), inteligencia (Nadia, 12:89) o interés (Nadia, 12:92).

Los participantes han expresado que los exámenes les permiten conocer acerca de sus alumnos su nivel de competencia, dominio, conocimiento, comprensión y grado de cumplimiento de objetivos. De igual manera, se mencionaron otros aspectos indirectos o poco relacionados con la validez, como motivación, actitud, estrategias y personalidad. No se profundizó en cuanto a estos últimos conceptos, pero dado que se relacionan poco con la evaluación de lengua podrían explorarse con mayor profundidad en el futuro y conocer más sobre qué aspecto de los exámenes les permite observar a los profesores esto.

No parece haber un consenso en cuanto a cuál es el objetivo principal en la evaluación del aprendizaje. La variedad de constructos definidos para ello, pueden ser resultado de una serie de factores diferentes relacionados con los profesores como sus antecedentes, formación, experiencia y creencias. No obstante, la competencia comunicativa es un constructo de tal complejidad que los objetivos de evaluación planteados por los profesores en las entrevistas tienen cabida en su conceptualización y que aplicado en los contextos específicos pueden dar como resultado instrumentos válidos. Dicho esto con la consideración de que la validez es una cualidad que se logra en grado y no en términos absolutos y que es resultado de una combinación de factores de muchos factores.

4.4.5 Dudas con respecto a los exámenes de lengua

Se preguntó a los maestros si habían dudado de algún examen y la razón de ello. La finalidad de esta pregunta fue para conocer sobre posibles prácticas que por sus resultados fuesen poco recomendables en el proceso de evaluación del aprendizaje de lengua.

Los aspectos del diseño y proceso de evaluación que fueron mencionados en las entrevistas fueron errores en el diseño del examen (Alondra, 6:175; Simón, 7:137), falta de revisión del examen por diversos motivos (Genaro, 10:146), como urgencia (Lorena, 1:178) y olvido

(Nora, 3:75); errores o poca claridad de las instrucciones; alta dificultad de los reactivos o de las tareas (Nora, 3:75; Camila, 4:48; Alondra, 6:180; Simón, 7:131); problemas en las tareas y problemas en los reactivos.

Los problemas que se encontraron con respecto a los reactivos fueron los siguientes: Ericka (2:69) comenta que los reactivos de opción múltiple no son confiables; Gina (5:36) mencionó que no contempló todas las posibles respuestas a algunas preguntas y que algunas otras son mal planteadas; Simón (7:135) mencionó ambigüedad en las respuestas de algunos reactivos; y Genaro (10:145) mencionó reactivos o preguntas confusas.

Los problemas con referencia a las tareas se dieron en que se incluyeron en el examen fueron aquellas que no se habían trabajado en el curso y que no eran familiares para los alumnos (Graciela, 8:57; Simón, 7:133). Lorena (1:184) explicó que por falta de tiempo solo tomó tareas del libro de texto.

Algunos profesores relacionaron las dudas acerca de los exámenes que surgieron en su experiencia con cualidades deseables como la validez y la confiabilidad. Lorena (1:188) mencionó que estas dudas no le permiten medir; mientras que Alondra (6:179) expresó que errores en el diseño afectan la confiabilidad. Otros profesores como Camila (4:46) y Genaro (10:144) manifestaron que debido a los malos resultados obtenidos por sus alumnos es que detectaron fallas en sus exámenes. Otros problemas de confiabilidad, aunque no manifestada explícitamente, obedecieron a aspectos relacionados con la aplicación, como problemas con el audio en la evaluación de la comprensión oral (Simón, 7:132), errores en el momento de imprimir el examen (Alondra, 6:177) y necesidad de dar indicaciones adicionales en el momento del examen (Simón, 7:139).

Esta pregunta también dejó ver algunas percepciones que tienen los profesores sobre los alumnos, como que contestan al azar las tareas de opción múltiple (Ericka, 2:70) o que finalizan demasiado pronto un examen (Nora, 3:74). De igual manera, percepciones de sí mismos surgieron en las respuestas como Virginia (11:124), quien confesó que en los inicios de su práctica hizo la reflexión sobre si en los exámenes ponía trampas a los alumnos.

Los aspectos que generan duda con respecto a los instrumentos de evaluación fueron relacionados con el diseño y proceso de evaluación, los reactivos, las tareas, los cuales convergen con principios de evaluación como validez y confiabilidad, además de algunos ajenos a los exámenes y más relacionados con los alumnos. Estas dudas entre los profesores tienen relación con sus reacciones acerca del uso de los exámenes como instrumentos de evaluación que se presentaron en este documento, renglones atrás. Al haber dudas en cuanto a la esencia de los exámenes, entiéndase validez y confiabilidad, se cuestiona su uso para evaluar.

Las dudas descritas por los profesores indican conciencia y reflexión hacia la propia práctica y han representado en varios de los casos expuestos oportunidades que han representado una mejoría en su desempeño profesional.

4.4.6. Autopercepción del profesor con respecto a la evaluación

Se le preguntó al profesor qué podía saberse acerca de ellos y su práctica docente al observar sus exámenes. Las respuestas fueron variadas y se presentan a continuación.

Lorena (1:209) hizo énfasis en que tiene “el perfil de lenguas”, probablemente haciendo referencia a aquellos profesores que tienen una formación distinta, además dio ejemplos de cómo el perfil profesional contribuye a que sus exámenes tengan bases y estructura. De igual manera, menciona la importancia de la validez de contenido y de realizar evaluaciones objetivas.

Ericka y Genaro mencionaron que sus exámenes reflejan algún gusto; en el caso de la primera, por la gramática y por cierto tipo de tareas tomadas de exámenes de certificación (Ericka, 2:124); y en el caso del segundo, hizo referencia a no ser balanceado, al tener preferencia por evaluar la producción escrita (Genaro, 10:171). Ericka también mencionó ser congruente al aplicar tareas adecuadas al nivel de lengua. Genaro (10:177) comenta que al revisar sus exámenes puede notarse que no se esforzó en buscar materiales adecuados, lo cual considera como “una pena”.

Gina y Simón consideran ser coherentes, aunque ella no profundizó en el tema, dejando la respuesta a sus alumnos. Por otro lado, Simón (7:170-171) explica que con coherencia se refiere a evaluar los contenidos estudiados en el curso, además de ser significativo, lo cual

de acuerdo con el profesor implica relevancia de las tareas y los contenidos. Gina (5:93) también hizo mención del término “responsable”, al igual que Alondra (6:233), quien agrega compromiso, planeación, e interés en su trabajo y en sus alumnos.

Algunas profesoras fueron muy concretas y descriptivas, se apoyaron en el uso de adjetivos para describirse, con base en sus instrumentos: Nora dice ser exigente y exagerada (3:216); Camila (4:105) organizada profesional y justa; Graciela (8:137) práctica; y Virginia (11:182) pragmática, sistemática y estricta (Virginia, 11:187). Nadia (11:125) se describe como “cuadrada”, lo que, como ella explica, implica que siempre sigue el mismo formato, no suele salirse de las reglas o de los estándares. Belinda (9:242) explica que sus exámenes muestran trabajo, objetividad, claridad y evidencia del aprendizaje.

Las respuestas a esta pregunta nos permiten observar cómo los profesores tienen, en su mayoría, una visión favorable de su propia práctica y una percepción de profesionales responsables, objetivos y comprometidos, entre varios calificativos de carácter positivo. Pocos fueron los casos con una percepción negativa, aunque con matices, como en el caso de las profesoras que se describieron como “cuadrada” o “desbalanceada”. En el caso de un profesor cuya opinión sí fue negativa en un caso extremo, se percibe que es debido a falta de experiencia y poca disposición al diseño de exámenes.

Las percepciones favorables con respecto a sí mismos como docentes es congruente con la que tienen de sus propios instrumentos, como se muestra a continuación.

4.4.7 Percepción de sus propios exámenes

La mayoría de los profesores tiene una opinión positiva acerca de los instrumentos elaborados por ellos mismos (Gina, 5:43; Graciela, 8:82; Nadia, 12:65). Algunos adoptan una postura moderada, como Nora (3:84), quien describe sus exámenes como “completos”, o Camila (4:59), quien los describe como “mejores” en relación con los que hacía previamente.

Esta postura es un indicio del desarrollo profesional de los profesores y muy probablemente de su literacidad en evaluación. La mirada retrospectiva genera una comparación de los instrumentos elaborados en el pasado y los presentes y con base en su experiencia en ese

inter ha acumulado una serie de elementos de carácter técnico, ético y lingüístico que le permiten dicho discernimiento.

Por otro lado, Genaro manifiesta una opinión negativa acerca de los instrumentos que ha desarrollado. En primer lugar, expresó no haber realizado muchos y cuando lo ha hecho han sido adaptaciones de exámenes existentes y sobre ellos comenta “dan un poco de lástima” (Genaro, 10:113). El profesor es de los que tienen menos experiencia de entre los participantes y en otros momentos de la entrevista expresó renuencia a evaluar con exámenes y admiración por instrumentos como los diseñados por Cambridge Assessment. Así que, esta opinión podría deberse a la poca experiencia en diseño de exámenes propios.

Estas percepciones positivas, tanto de su propia práctica, como de sus instrumentos de evaluación, pudiesen parecer poco objetivas. No obstante, a lo largo de las entrevistas y de los análisis realizados es posible percibir reflexión y autocrítica por parte de los participantes, como se pudo apreciar incisos arriba en este mismo apartado.

4.4.8 Reflexión sobre el impacto de sus instrumentos en los alumnos

Se preguntó a los profesores si habían reflexionado sobre el impacto que podría tener el examen en la vida de los alumnos. La mitad de los profesores respondió afirmativamente y la mitad negativa. Algunos profesores dieron un ejemplo de una situación donde se dio ese impacto (Nora; Camila; Genaro). Genaro (10:166) y Camila (4:93) comentaron que reprobar para el alumno hubiese significado baja de su sistema escolar

Gina y Alondra comentaron que un mal diseño del examen tiene impacto en los alumnos. Belinda (9:222) comentó que no había realizado esa reflexión debido a que se enfoca en el aprendizaje y en el examen como medio de aprendizaje.

Esta pregunta está relacionada estrechamente con el código que se denominaría alteridad y es evidencia de conducta ética por parte de los profesores y, de igual manera, del cuidado que tienen para los alumnos. No obstante, es notable el hecho que la mitad de ellos no haya hecho la reflexión acerca del impacto de sus instrumentos de evaluación en sus alumnos, hecho que pudiese deberse a experiencia, falta de desarrollo de literacidad en evaluación o a una centralización por parte de los profesores en su propia práctica y persona.

Conclusiones a la fase cualitativa de análisis

Se realizó el análisis de la información obtenida de 12 entrevistas, en tres etapas, de acuerdo con la metodología de la teoría fundamentada: Codificación abierta, codificación axial y codificación selectiva.

En la codificación abierta se asignaron códigos a todos los segmentos de discurso en las entrevistas y se encontraron relaciones entre los códigos. Estas relaciones se establecieron siguiendo dos criterios que se denominaron relación por incidencia y relación teórica. Se pudo observar el papel central que juegan diversos aspectos de confiabilidad y validez en estas relaciones. Estos códigos se relacionan, por incidencia, con varios códigos de la dimensión ética, como alteridad, cuidado del alumno, reflexión, exclusión e inclusión. El código habilidades, de la dimensión lingüística, entre los principales códigos, es el único que no pertenece a las dimensiones ética y lingüística.

En la codificación axial se agruparon los códigos de acuerdo con características en común. Estas agrupaciones resultaron en cinco categorías a las cuales se les llamó dimensiones que son: Lingüística, técnica, ética, personal y docente. Dentro de la dimensión ética se encontraron dos subgrupos principales, a los que se denominó subdimensión valorativa y subdimensión interpersonal.

La codificación selectiva tuvo como resultado un código central donde se integran las dimensiones, que es la responsabilidad profesional. Se propone como definición del término a la capacidad de un profesional de la enseñanza de lengua para conjugar los factores técnicos, éticos, lingüísticos, docentes y personales en la toma de decisiones en su práctica evaluativa.

En el capítulo se presentó el análisis de cada dimensión y sus códigos más representativos, lo que se determinó por su presencia en el discurso. Se dio cuenta de las relaciones que se establecieron con otros códigos y en algunos casos marcadores discursivos que indicaban la pauta a seguir para la asignación de determinados códigos. Cada dimensión y código se ejemplificaron con citas extraídas del discurso de los profesores participantes.

Se realizó el análisis del discurso de los profesores para cada pregunta de la entrevista y así lograr una visión temática que permitiera al investigador conocer las percepciones de los

profesores en relación con ciertos temas en particular. Dicha visión se diluía en el proceso metodológico de la teoría fundamentada, por lo que se decidió realizar este análisis que permitiera una descripción de dichas percepciones de manera puntual.

El análisis cualitativo presentado es evidencia de la conformación del discurso de los profesores acerca de su práctica en diseño de exámenes y evaluativa y una deconstrucción de estas, cumpliendo así con el objetivo general. Los códigos y las dimensiones describen cómo se estructura el discurso alrededor de un eje que es su propia responsabilidad profesional. De igual manera, se muestra su interacción envolvente, la complejidad que solo sale a la luz a través de la introspección de los participantes, de sus reflexiones y de la conjugación de todos los factores en el contexto y personas en particular.

En el siguiente capítulo se profundizará en los hallazgos de los análisis, el desarrollo de la teoría originada, su relación con la teoría presentada en capítulos previos y posibles repercusiones en la práctica evaluativa del profesor de lenguas.

DISCUSIÓN DE RESULTADOS

En este apartado se discuten los hallazgos en relación con las preguntas de investigación y objetivos planteados, así como de los que surgieron de los análisis de las fases de análisis cuantitativa y cualitativa. En estas fases se obtuvieron categorías, factores en el caso del análisis cuantitativo y códigos y dimensiones en el caso del análisis cualitativo. En seguida se comparan estos resultados y se estudian a la luz del pensamiento complejo. El resultado más relevante se da en la concepción de un término que engloba varios factores de la práctica evaluativa y de diseño de exámenes en un contexto complejo y que explica las razones que llevan a los profesores a tomar decisiones con respecto a esta práctica: la responsabilidad profesional del profesor en la evaluación del aprendizaje. Además de la estructuración de la propuesta de la teoría mencionada, se presentan las conclusiones más relevantes del estudio y su relación con la teoría expuesta en capítulos anteriores.

La primera pregunta planteada fue acerca de conocer las prácticas evaluativas de los profesores de la Universidad. La encuesta realizada en la fase cuantitativa arrojaría datos sobre la importancia de las prácticas relacionadas con los aspectos éticos de la evaluación y la consciencia acerca del impacto de los exámenes en los alumnos. Los resultados de la fase cuantitativa también indicaron que las prácticas evaluativas de los profesores se acercan a las prácticas deseables, que fueron determinadas con base en la literatura y en entrevistas previas a profesores con experiencia en enseñanza de lengua y evaluación.

Estos hallazgos dieron lugar a nuevas interrogantes, sobre cómo construye el profesor su práctica profesional, sobre cómo toma decisiones acerca de las evaluaciones y del diseño de exámenes en particular. De igual manera surgieron dudas acerca de porqué los profesores tenían una visión de su propia práctica y que no era congruente con la visión de otros participantes del proceso de evaluación, como directivos y particularmente, alumnos. La construcción de la práctica evaluativa en palabras del propio profesor se volvió el tema eje de la investigación.

De este modo, se dio a la segunda fase del proyecto, la cualitativa, el estatus dominante del estudio. La sola descripción de las prácticas no eran suficientes para entender qué había

detrás de las decisiones del profesor, entonces se encontró el fundamento epistemológico que permitiría adoptar una visión más completa y entender un fenómeno con tantos factores tan diversos como la evaluación del aprendizaje de lengua. Se encontró en la teoría de la complejidad (Morin, 1990) el paradigma epistemológico que daría cabida a la visión necesaria para entender la problemática y elegir el rumbo metodológico. Se adoptó la teoría fundamentada como la tradición del enfoque cualitativo que se adaptaba mejor a las necesidades metodológicas del estudio.

Con base en la metodología de la teoría fundamentada y el análisis del discurso fue posible determinar los factores que estructuran la práctica evaluativa a los cuáles se les agruparía y denominaría dimensiones. Los datos obtenidos en la fase cuantitativa se utilizarían para realizar un análisis factorial que daría validez a los resultados de la fase cualitativa, al encontrarse una correspondencia entre las dimensiones de la fase cualitativa y los factores de la fase cuantitativa.

En las dimensiones de la práctica evaluativa y sus intrincadas interconexiones se encontró una semejanza entre el proceso evaluativo y un sistema complejo con todas sus características y así se busca entender. Más sobre esta propuesta en algunos incisos más adelante en esta misma sección.

Las dimensiones parecían conformar la base para un nivel superior de decisión, que conjugara estas y conformara una estructura de discurso sobre la práctica y base para las decisiones del profesor, lo que resultaría en la propuesta de la responsabilidad profesional, un término que, además de cubrir las dimensiones, da cuenta por sí mismo de cómo funciona el sistema de decisiones en la práctica evaluativa, con las dimensiones ética y técnica entrelazadas sólidamente apoyándose de la dimensión lingüística para conformar dicho sistema. La fase cualitativa de este estudio permitió conocer cómo se interrelacionan las dimensiones y cómo se estructura la responsabilidad profesional a lo largo del discurso.

Los incisos siguientes dan cuenta pormenorizada de estos hallazgos y otras reflexiones por parte del investigador y que surgieron a la luz de los análisis y resultados de las fases de estudio.

5.1 Triangulación de resultados

En este apartado se da cuenta de las respuestas a tres preguntas de investigación planteadas acerca de los factores que sirven como base para la toma de decisiones del profesor en su práctica evaluativa; sobre cómo el profesor construye su responsabilidad profesional en su discurso; y acerca de cómo se da la interrelación de las dimensiones de la RP en el discurso del profesor.

Adicionalmente, y con el propósito de validar los resultados obtenidos en la fase cualitativa, se llevó a cabo el análisis estadístico que nos permitiera realizar una triangulación entre los resultados de esta fase y la cuantitativa. Cabe mencionar que el tamaño de la muestra no permite aseveraciones o generalizaciones significativas, sin embargo, dado que se trata de un estudio de caso y que el objetivo es apoyar el análisis cualitativo que nos permite generar la teoría para explicar el fenómeno. Al relacionar los factores encontrados en el análisis factorial con las dimensiones resultantes del análisis cualitativo se encontraron coincidencias que se describen a continuación.

El análisis factorial de los resultados obtenidos con la encuesta aplicada a la muestra de los profesores de lengua inglesa y francesa de la Universidad mostró seis factores de análisis que surgen de los resultados en la encuesta. Los factores son: Valores e impacto, responsabilidad en el proceso de evaluación, aspectos técnicos, propósitos y objetivos del examen, utilidad del examen y preparación del examen. Lo aquí expuesto contribuye a responder la pregunta sobre los factores que tienen efecto en la toma de decisiones de los profesores sobre su práctica evaluativa.

El análisis cualitativo resultó en cinco dimensiones: Dimensión técnica, dimensión lingüística, dimensión ética, dimensión personal y dimensión docente. Cada dimensión consiste en un numeroso despliegue de códigos, siendo los más importantes: confiabilidad, validez, validez de contenido y tareas, de la dimensión técnica; alteridad, inclusión, exclusión, reflexión y cuidado del alumno, como elementos de la dimensión ética; y habilidades, código que corresponde a la dimensión lingüística.

Se relacionaron los factores resultantes del análisis factorial con las dimensiones del análisis cualitativo y se determinaron las siguientes relaciones:

Tabla 5.1*Relación entre factores y dimensiones*

Factores resultantes del análisis cuantitativo	Dimensiones resultantes del análisis cualitativo
<ul style="list-style-type: none"> • Valores e impacto • Responsabilidad en el proceso de diseño de examen 	Dimensión ética / técnica
<ul style="list-style-type: none"> • Aspectos técnicos • Utilidad • Preparación del examen 	Dimensión técnica
<ul style="list-style-type: none"> • Propósito y objetivos del examen 	Dimensión lingüística / técnica

En primera instancia, el factor Valores e impacto describe acciones relacionadas con la congruencia entre los valores del profesor con los educativos e institucionales y la conciencia sobre las consecuencias derivadas de los resultados de los exámenes. Este factor coincide con los aspectos éticos encontrados en el discurso de los profesores, especialmente la alteridad, el impacto y cuidado del alumno; también se relaciona con la dimensión técnica, particularmente con la validez, ya que los reactivos 57 y 59 del cuestionario hacen referencia a las interpretaciones que se hacen de los resultados, sus consecuencias y explícitamente la validez del instrumento. El factor Responsabilidad en el proceso de diseño de examen integra reactivos del cuestionario que describen la conciencia del evaluador sobre los aspectos técnicos necesarios para el diseño del instrumento, así como de su responsabilidad en el proceso, mismos aspectos que se consideraron dentro de la dimensión técnica.

Tres de los factores resultantes del análisis factorial, aspectos técnicos, utilidad del examen y preparación del examen, se relacionan directamente con la dimensión técnica que fue resultado del análisis cualitativo. Esta dimensión contempla los aspectos técnicos del proceso de diseño, aplicación y calificación de instrumentos de evaluación del aprendizaje de lenguas, como validez, confiabilidad y practicidad. En el análisis factorial se obtuvo el factor Utilidad, que se relaciona con el concepto *test usefulness* (Bachman y Palmer, 1996),

que engloba estos mismos conceptos. El factor 3, Aspectos técnicos, contempla conceptos como conocimiento y procedimientos técnicos, las tareas en un examen, tablas de especificaciones y otros relacionados con la confiabilidad de un instrumento. El factor 6, Propósito y objetivos del examen, también se relaciona con aspectos técnicos, en este caso de dos aspectos de la preparación de un examen: el piloteo de los exámenes y el entrenamiento de los profesores.

El factor 4, Propósito y objetivos del examen, se relaciona con la validez de los instrumentos de evaluación, en el entendido de que estos cumplen su propósito y objetivos de evaluación. La validez se ha considerado como un aspecto de la dimensión técnica, pero también se relaciona con la dimensión lingüística. En general, los exámenes de aprendizaje de lengua deben medir la competencia comunicativa, si cumplen dicho propósito se considerarán válidos. De aquí que el factor se relacione con estas dos dimensiones.

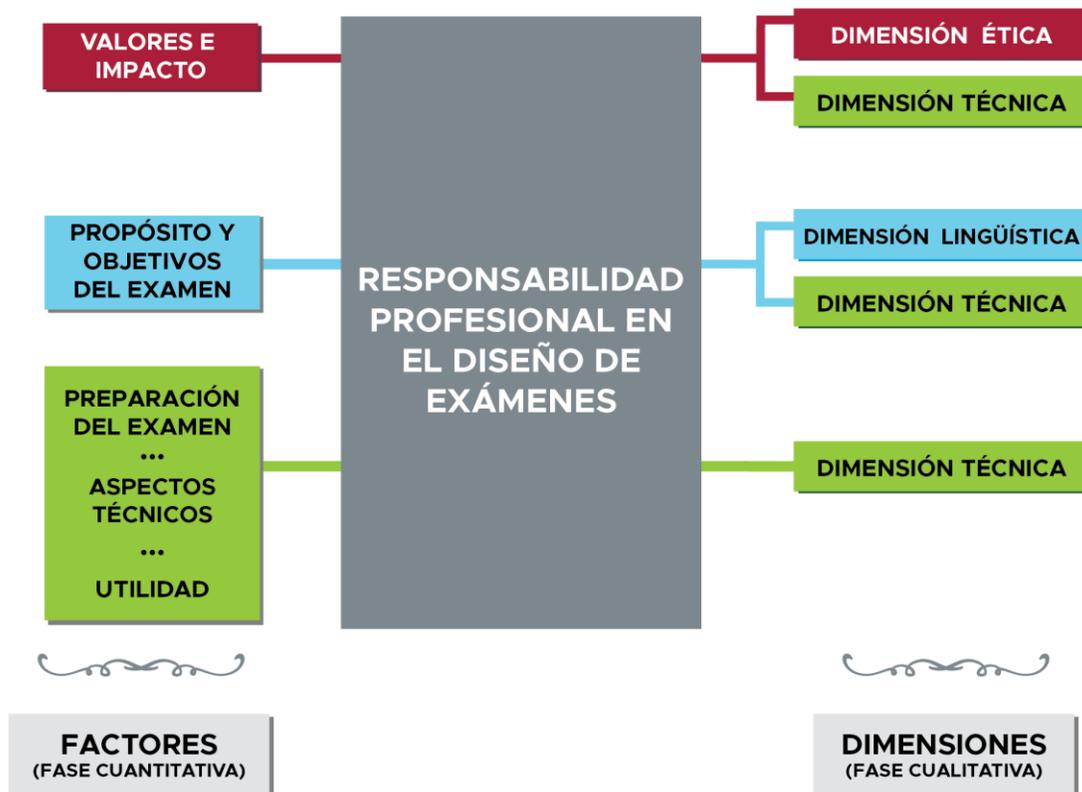
Con base en lo anterior, se considera que es posible establecer que existe una relación entre los resultados de la fase cuantitativa y los resultados de la fase cualitativa que podría sustentar estos últimos. La coincidencia entre los factores y las dimensiones principales se da en elementos en común como son responsabilidad, validez y confiabilidad, principalmente. Los factores en la fase cuantitativa indican que en el caso de los profesores sujetos de esta investigación conceptos como valores, responsabilidad, validez y confiabilidad son trascendentes para su práctica evaluativa; lo que coincide con lo encontrado en el discurso de los profesores entrevistados, cuya descripción de su práctica puede concentrarse en tres dimensiones principales: la dimensión ética, técnica y lingüística. Los mismos conceptos resultantes del análisis factorial integran códigos de dichas dimensiones.

En el siguiente diagrama (Figura 5.1) se muestra la relación entre factores y dimensiones. En primera instancia, los valores educativos, institucionales y personales, junto con la conciencia acerca del impacto de los resultados de una evaluación llevan al establecimiento de un propósito y de los objetivos de un examen. Esto lleva a la preparación del examen y aplicación de los aspectos técnicos y de búsqueda de utilidad del examen, es decir aquellos relacionados con la validez, confiabilidad y practicidad del instrumento. Todo lo anterior se apoya y a la vez conforma un sistema de responsabilidad en el proceso de diseño que se

basa en las dimensiones ética, técnica y lingüística, que se encontraron en el discurso de los profesores sobre evaluación y diseño de exámenes de aprendizaje de lengua.

Figura 5.1

Factores de decisión y dimensiones de la responsabilidad profesional.



Nota: Elaboración propia.

Con base en los resultados de las encuestas y el análisis del discurso de los profesores se ha integrado una estructura inicial que indica cómo podría construirse la práctica evaluativa del profesor de lengua. Los factores resultantes del análisis cuantitativo y las dimensiones y códigos que resultaron del análisis cualitativo están entrelazados ya que los componentes en sus unidades, como son los códigos, son los mismos. Esta constitución del proceso puede entenderse con la perspectiva de la complejidad, como se explicará en el siguiente apartado.

Desde la perspectiva del paradigma tradicional de la ciencia positivista, sería necesario aportar más datos que resulten en la significatividad estadística necesaria para describir una

práctica evaluativa general de los profesores de lengua de la universidad, puesto que los datos aportados por esta investigación no serían suficientes, debido al tamaño de la muestra. Sin embargo, la postura epistemológica bajo la cual se realizó este estudio demanda una visión más integral, en el entendido de una problemática compleja.

5.2 Los resultados vistos desde la perspectiva de la complejidad

En contraste a lo anterior, desde la perspectiva de la complejidad, la metodología con que se ha estudiado el caso nos ha permitido acercarnos a comprender el entramado con que se construye la práctica evaluativa. Los factores encontrados en la fase cuantitativa asoman el hilo conductor que nos lleva a la construcción de la teoría desarrollada. Con base en la metodología de la teoría fundamentada, aplicada en el estudio cualitativo de esta investigación, se han logrado identificar las dimensiones que conjuntan los códigos del discurso de los profesores. Estos códigos son los componentes individuales del discurso y de la práctica evaluativos, se integran en los segmentos del discurso, construyendo la descripción de las prácticas, creencias, ideologías y sentires de los docentes participantes en el estudio.

Los códigos, a su vez, componentes de las divisiones, tienen presencia en sus equivalentes en la fase cuantitativa; de esta manera, conceptos como responsabilidad, validez, confiabilidad, practicidad son una constante en ambos enfoques de la investigación, códigos y factores que se relacionan, repiten y revelan componentes esenciales, predominantes en la práctica evaluativa de los docentes. No obstante, la reafirmación de dichos conceptos como esenciales en la práctica no es un indicio de delimitación o exclusión de elementos menores. Con base en los valores epistémicos postulados por Romero Pérez (2003), los conceptos teóricos existentes se han replanteado, ubicándoles en su contexto particular, en la acción, (Romero Pérez, 2003), se han buscado nuevas aplicaciones prácticas y se han encontrado las relaciones que unifican el conocimiento en varios niveles de interconexión, como plantea Prigogine (1995).

El estudio cualitativo nos ha permitido visualizar una compleja estructura del discurso que soporta las decisiones en la práctica evaluativa. La complejidad de esta estructura se da en lo numeroso de los códigos que la componen, así como en las, también numerosas, relaciones que surgen del análisis discursivo. Al agrupar los códigos en dimensiones no se

está incurriendo en la simplificación o generalización, esperada del método científico (Morin, 1990); sino que, a la luz de la complejidad, estas dimensiones se perciben como sistemas que interactúan entre sí, conformando uno nuevo, igualmente complejo, autorregulado, con interacción con agentes externos, que se renueva y adapta a los cambios, es lo que guía, regula, organiza y desarrolla la práctica evaluativa del docente.

Una vez concluido este estudio, se ha logrado conforme a los argumentos planteados por De Almeida (2008) para entender la complejidad, una visión unitaria, más no simplificante, de la práctica docente. Se observa un fenómeno complejo, con cierto grado de incertidumbre e imprevisibilidad, no lineal, auto-organizado, en evolución, dependiente y autónomo a la vez, con emergencias implicadas y con necesidad de equilibrio. Es posible percibir todo lo anterior en el discurso de los profesores, la variedad de perfiles, a pesar de ser todos profesores de la Universidad, muchos de ellos egresados de esta misma, repercute en decisiones tan diversas que le dan el carácter de sistema incierto, imprevisible, regulado por las relaciones laborales e institucionales, pero a la vez independiente en las prácticas individuales de los profesores. La práctica evaluativa es la misma, pero a la vez es diferente en cada individuo, en cada profesor que es una unidad, un ente autónomo que, aunque dirigido por dichas relaciones, toma decisiones con base en una diversidad de factores, representados en los códigos.

Los grupos de profesores se organizan de manera independiente a la administración escolar y a la vez cada profesor organiza su propia forma de trabajar, todos como unidades autopoieticas (Maturana y Varela, 1994). El profesor se prepara, evoluciona con cursos de capacitación, con retroalimentación de otros colegas; de igual forma que las instituciones capacitan a sus profesores y al hacerlo evoluciona el conjunto. Tómese como ejemplo a Lorena y Alondra, las profesoras más experimentadas, y a Graciela, una de las profesoras con menos experiencia. Las primeras muestran en su discurso un alto contenido de aspectos técnicos de la evaluación de lengua. No obstante, es posible notar diferencias importantes, entre ellas en otros aspectos como el lingüístico, en el que Alondra se muestra como la más baja entre los participantes con contenido de esta dimensión en su discurso. Graciela, la segunda más baja en este aspecto muestra un alto contenido de factores éticos en su discurso. Podría describirse a Lorena y Alondra como unas profesoras que diseñan instrumentos con un alto grado de tecnicidad, aunque la segunda mostraría un enfoque

mucho menor en el objeto de la evaluación que Lorena. En contraste, podría asumirse que la práctica evaluativa de Graciela tiene una fuerte carga ética, aunque se deja un poco de lado el aspecto lingüístico. Es necesario puntualizar que, estos supuestos todavía deben confirmarse con análisis de los instrumentos de evaluación elaborados por las profesoras participantes.

La impredecibilidad de los sistemas complejos se da debido a la estrecha conexión entre sus numerosos factores. Un cambio en uno de ellos significa una recomposición en varios otros. En el caso de la práctica evaluativa, alguna modificación en la políticas o instrucciones por parte de la institución dirigente, alguna nueva experiencia por parte del profesor, personal, académica, laboral, o algún evento de aprendizaje por parte del alumno, o simplemente nuevos alumnos, con características diferentes a las previstas por el profesor, todos estos factores dan lugar a modificaciones en el funcionamiento del sistema y que son imposibles de predecir. De acuerdo con Larsen-Freeman (2011) el uso de la teoría de la complejidad no es para fines predictivos, sino explicativos, para entender el pasado; el futuro en los sistemas complejos es simplemente impredecible.

El sistema se adapta, se equilibra, porque sus unidades así lo hacen, como Alondra al adaptarse a los requerimientos específicos del proceso evaluativo; Genaro al reflexionar en el momento del discurso sobre su propia práctica, o Erika al modificar su práctica para cumplir con los requerimientos de su coordinadora.

El estudio de este caso contempla otras variables que contribuyen a su complejidad además de las características individuales de los profesores, como el contexto, con que el sistema interactúa y que tienen injerencia en sus decisiones, a manera de autorregulación. Las autoridades de los espacios universitarios donde se desempeñan los profesores; las necesidades particulares de los alumnos, de igual manera todos con diferentes perfiles; o las condiciones laborales de los profesores, son todas situaciones ajenas al proceso de evaluación del aprendizaje, elementos del contexto que afectan la práctica del docente. Esta se ajusta a las condicionantes contextuales por medio de las decisiones del profesor.

5.2.1 La evaluación del aprendizaje de lengua como un sistema

Lo expuesto en el inciso anterior permite entender la evaluación del aprendizaje de lenguas que se lleva a cabo en el salón de clase como un sistema complejo, donde se conjugan factores externos al aula, como las instituciones y todo lo que representan, su administración, directivos y políticas. Otro de los factores externos que conforman el contexto son los que conforman la sociedad, como padres de familia y empleadores; y factores internos al aula, y que son parte del profesor y del alumno.

En el caso del profesor, su conocimiento técnico sobre evaluación, su conocimiento lingüístico, experiencia, y ética personal y profesional construyen en el profesor un bagaje que junto con los factores externos, principalmente la institución como empleadora y legisladora, influyen en las decisiones que debe tomar el profesor relacionadas con la evaluación, más específicamente con el diseño, aplicación y calificación de exámenes o instrumentos.

Esta evaluación se aplica al alumno, quien, a su vez, con influencia de otra serie de factores innumerables, obtiene un resultado que es interpretado por el profesor y por la institución. Con base en esta interpretación del resultado estos toman decisiones que repercuten en el alumno. La información es procesada nuevamente tanto por el profesor, como por la institución de manera que se toman nuevas decisiones que nuevamente dan inicio al ciclo, que puede verse ilustrado en la figura 5.2.

El ciclo descrito es común para todos los casos, pero diferente en cada uno de ellos. Cada uno de los participantes, es decir, las instituciones, los profesores y los alumnos, son las unidades autónomas que se han descrito en este trabajo, que se autorregulan y que se adaptan a cada variante de manera que el sistema siga funcionando.

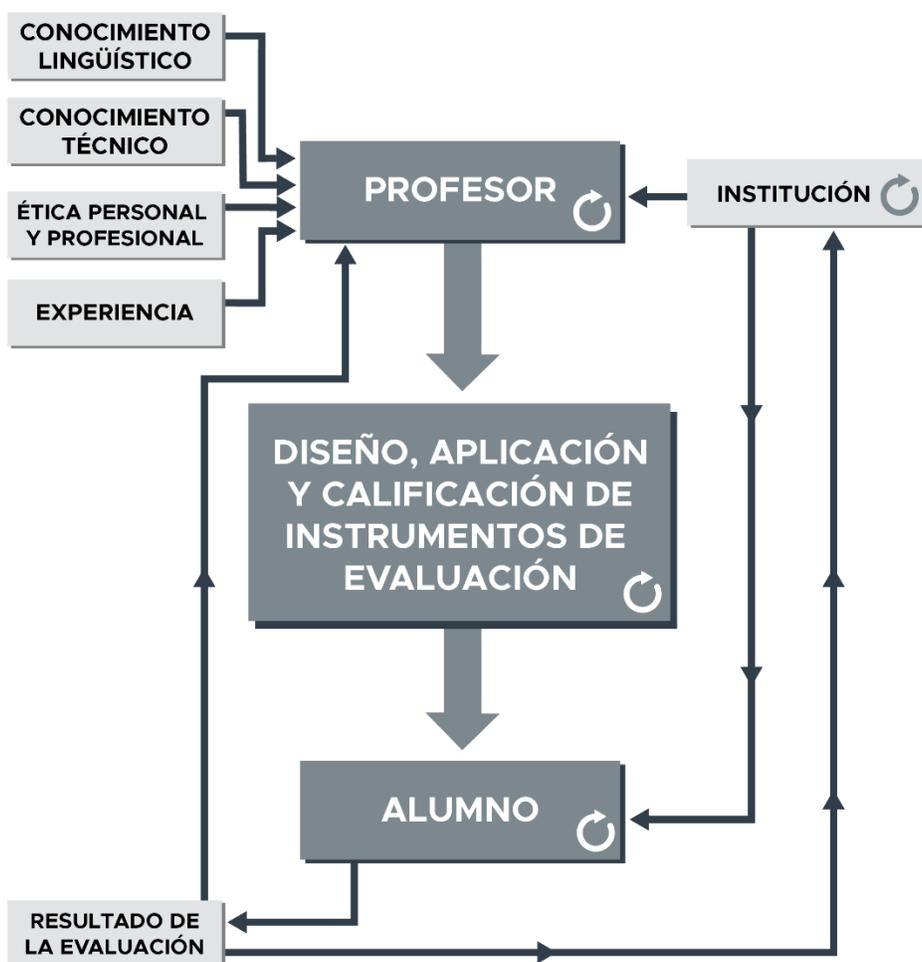
El profesor, que es el objeto de esta investigación, debe adaptarse a los cambios en cuanto a políticas, regulaciones y limitaciones de índole académica y administrativa por parte de las instituciones. De igual manera nuevas experiencias, entrenamiento, reflexiones, intercambios de ideas y experiencias en colaboraciones con colegas influyen en cambios y nuevas adaptaciones por parte del docente. Estas adaptaciones también se realizan con base

en el alumno, en el individuo, no solo en el genérico, por lo que cada alumno debería representar una adecuación del sistema, tarea sin duda compleja.

La práctica responsable debe guiar la conducta y la práctica del docente, adecuarla de manera que los cambios contextuales y en los individuos converja en el aprendizaje del alumno, a su cuidado y a la construcción del profesionista y profesional.

Figura 5.2

La evaluación del aprendizaje de lenguas como un sistema



Nota: Elaboración propia.

5.3 La responsabilidad profesional del profesor en la evaluación del aprendizaje de lengua

Como se ha visto hasta este momento, la práctica evaluativa de los profesores de lengua de la universidad obedece a numerosos factores que se relacionan entre sí, que se integran, que llevan al profesor a la toma de decisiones y que tienen un impacto en todo el sistema. Es así como se ha conformado el concepto rector de la práctica evaluativa, al que se ha denominado *responsabilidad profesional* del profesor de lengua en la evaluación del aprendizaje y que se propone como un sistema de toma de decisiones del profesor para la evaluación del aprendizaje de lenguas.

La responsabilidad profesional es el resultado de la codificación selectiva del proceso de análisis con base en la teoría fundamentada y para este caso de estudio, se entiende como la capacidad de un profesional de la enseñanza de lengua para conjugar los factores técnicos, éticos, lingüísticos, docentes y personales en la toma de decisiones relacionadas con su práctica evaluativa y que opera en esta última como un sistema.

Este sistema rector consiste de tres dimensiones o subsistemas principales que son la ética, técnica y lingüística, así como de dos dimensiones menores, al menos en el discurso, que son la docente y la personal. Estas dimensiones se componen de códigos o factores, que se perciben como las unidades del proceso. Al ubicar las dimensiones como los grandes grupos de factores y sus relaciones se conforma una nueva unidad, que guía la práctica. Es decir, las decisiones que toman los profesores con respecto a la evaluación del aprendizaje de sus alumnos tienen sus bases en estas dimensiones, lo cual es congruente con lo declarado por Hamp-Lyons (2000).

La responsabilidad profesional guía la práctica evaluativa del profesor de lengua; sus decisiones son dirigidas por esta responsabilidad integrada de factores éticos, técnicos y lingüísticos. El proceso de evaluación del aprendizaje de lengua se integra de varios elementos y participantes: las condiciones y actores sociales, como empleadores y padres de familia; la institución educativa, con su misión, legislación, infraestructura, políticas y limitaciones presupuestales; los estudiantes, con sus necesidades particulares, perfiles, intereses, actitudes; el profesor de lengua, con sus antecedentes personales, educativos y

profesionales, intereses, principios, motivaciones y limitaciones. Todos ellos se conjugan para crear una situación única en cada evaluación y en cada instrumento.

El profesor es el eje y actor más importante en el proceso, utiliza su conocimiento técnico, lingüístico y principios éticos para diseñar un instrumento de evaluación. El contexto, donde se integran las instituciones, los actores sociales y los alumnos, influye directamente en el trabajo del profesor. Éste toma en consideración estos factores y combina con los propios para diseñar una evaluación y un instrumento para llevarla a cabo. Los instrumentos tendrán distintos grados de validez, resultado del logro de cualidades de utilidad, en otras palabras, del logro de validez del constructo, validez de contenido, confiabilidad, practicidad, autenticidad e interactividad (Bachman y Palmer, 1996; Brown y Abeywickrama, 2018).

En relación a lo anterior y como pudo observarse en el análisis cualitativo de este estudio los códigos o factores más importantes, por su frecuencia en el discurso y sus relaciones o conexiones con otros códigos, son validez, confiabilidad, validez de contenido y tareas, de la dimensión técnica; alteridad, cuidado del alumno, inclusión, exclusión y reflexión de la dimensión ética, específicamente en la denominada subdimensión interpersonal; y habilidades, de la dimensión lingüística; y dentro de estos mismos códigos vale resaltar la importancia de la confiabilidad y de la validez. La confiabilidad de un examen toma un papel central, lo que se debe a las conexiones con varios códigos, tanto de su propia dimensión, la dimensión técnica, como de la dimensión lingüística y dimensión ética, y que se pueden distinguir gráficamente en la figura 4.1. Teóricamente, el factor más importante debería ser la validez (Bachman y Palmer, 1996; Fulcher y Davidson, 2007), cualidad que debe asegurarse en todos los instrumentos de evaluación, sin embargo, la confiabilidad se relaciona con varios factores más que la propia validez y ocupa un lugar preponderante en la realidad de la práctica de los participantes en el estudio. Cabe aclarar que esto no significa que la confiabilidad sea más importante que la validez. La confiabilidad se percibe como un componente vital en el logro de validez en los instrumentos, la aplicación de mecanismos prácticos que coadyuvan en este logro.

Confiabilidad y validez son conceptos técnicos del diseño de exámenes y de evaluación que son aplicados en la práctica docente por los profesores entrevistados, aunque no siempre

con plena conciencia de su uso, conceptualización o, incluso, existencia. En este trabajo se ha expuesto cómo se relacionan con gran parte del discurso de los participantes y su relevancia en sus decisiones, de igual manera, cómo interactúan con las dimensiones ética y lingüística. El discurso de los profesores evidencia también la importancia del aspecto ético e interpersonal en esas mismas decisiones. Los otros dos aspectos más importantes en el discurso fueron las tareas a realizar por parte de un alumno en un examen como medio para recolectar información acerca de su competencia comunicativa y las habilidades a ser evaluadas, ya sea en general o en específico.

La validez se relaciona estrechamente con las habilidades y constructo a ser evaluados, con el contenido de los exámenes y con la propia confiabilidad, que cabe mencionar, es un requisito indispensable para la consecución de la misma validez. Además de ésta, la confiabilidad se encontró en relación con numerosos factores de la dimensión ética, como son exclusión, inclusión, alteridad, reflexión y de la dimensión técnica, con el código tareas. La importante conexión de la confiabilidad con estos conceptos de la dimensión ética-interpersonal podría indicar una conciencia de parte del evaluador de los papeles que juegan los participantes en el proceso de evaluación. Al referirse a otros (alteridad, inclusión), a sí mismo (exclusión), mientras se refiere a las acciones relacionadas con la confiabilidad, nos da un indicio de ello. El hecho que las prácticas relacionadas con el logro de la confiabilidad generen reflexión en los profesores, también indica una influencia de la dimensión ética en los aspectos técnicos de la práctica evaluativa.

Es por lo anterior que, es posible argumentar que la responsabilidad profesional del profesor de lengua en la evaluación del aprendizaje obedece a una variedad de factores de las tres dimensiones mencionadas, sin embargo, ciertos aspectos o factores de estas son más trascendentes, como los mencionados en los párrafos anteriores. Por consiguiente, se percibe a esta responsabilidad profesional como el factor de decisión en la práctica de los profesores; el componente en el que se interrelacionan las dimensiones descritas a manera de un sistema organiza, autorregula y adapta a su contexto.

5.4 Profesionalismo y aspectos éticos de la práctica evaluativa

La ética se entiende como se ha planteado en el capítulo II de este trabajo, con base en la ética compleja y la religación de Morin (2006), la ética profesional de Ibarra (2005), la

ética del docente universitario (Román, 2003; Martínez Navarro, 2010) y la responsabilidad (Davies, 2010). Dado lo anterior, el ejercicio profesional del docente de lengua implica un desempeño por el bienestar del individuo, la sociedad y la especie (Morin, 2006), con responsabilidad y cuidado del alumno, en su relación desigual con el profesor.

En ambas fases de la investigación, se ha encontrado que los aspectos éticos tienen relevancia en la práctica evaluativa de los profesores. En la fase cualitativa, en la codificación axial, surgieron varios códigos que pudieron ser agrupados en lo que posteriormente se denominaría Dimensión Ética, éstos fueron catalogados en dos subdimensiones, la subdimensión personal y la subdimensión valorativa. Se consideró denominarlas así, en el primer caso porque, como se explicó previamente, la subdimensión personal es evidencia de cómo se estructuran y cómo funcionan los mecanismos de las relaciones entre los individuos que participan en el proceso de evaluación del aprendizaje; y la segunda, la subdimensión valorativa, porque engloba aquellas citas que son evidencia de juicios emitidos por los entrevistados y que se encaminan a una regulación por parte del profesor, de su conducta o acciones o las de los alumnos y colegas. Ambas subdimensiones se explican a continuación.

5.4.1 Ética interpersonal y valorativa

En su dimensión interpersonal la ética en la evaluación, a partir de los resultados, se entiende de la siguiente manera:

A través de la alteridad, esa consideración del Otro y de su experiencia, el profesor encuentra explicaciones, causas y razones para su práctica evaluativa. Esto repercute en un cuidado del alumno generado por el interés en el aprendizaje del alumno, sentimientos de responsabilidad, tolerancia, objetividad, e interés o atención del docente en la formación personal del alumno, adicionalmente a su formación académica.

La ética se manifiesta de igual manera a través de los principios de exclusión e inclusión. La exclusión es una centralidad, una predominancia del Yo, que busca distinguir al sujeto de otros, en este caso al profesor de otros profesores y de sus alumnos. Por otro lado, el principio de inclusión, que es el Nosotros, el principio que genera comunidad entre el profesor y su alumno, entre él y sus colegas profesores, él y su institución y que permite al

sujeto generalizar en primera persona, haciendo partícipes de sus decisiones a todos estos agentes.

Otro factor que influye en la ética del profesor es el impacto que tienen los exámenes tanto en el profesor, como en los alumnos y puede ser positivo o negativo. El impacto también comprende al ámbito personal, además del académico.

Es a través de la reflexión y del diálogo que el profesor logra permear toda su experiencia y sus relaciones con los diversos agentes participantes en la práctica del proceso evaluativo. Esta reflexión le genera duda, le hace considerar al alumno y al impacto de sus evaluaciones; y el diálogo que en ocasiones es real y en otras hipotético, le lleva a la negociación, le obliga o motiva a la retroalimentación y a la recomendación.

En resumen, la ética interpersonal del docente se rige por los siguientes factores:

- La alteridad
- Principios de exclusión e inclusión
- El impacto
- La reflexión
- El diálogo

El profesor refleja su ética interpersonal en su práctica evaluativa a través de estos principios; las relaciones que el profesor construye y el posicionamiento que toma con respecto a sí mismo y con terceros, son la estructura de esta ética que los profesionales de la docencia de lengua ejercen. Esta visión de la ética es congruente con la ética compleja de Morin (2006), particularmente en el concepto de religación ética, ya que se percibe el interés del profesor, el sujeto en el otro, su alumno y en la sociedad.

La dimensión ética en la práctica evaluativa de los profesores muestra una faceta valorativa, en la que se emiten juicios o valoraciones que dejan ver cómo se estructura este aspecto ético en su discurso. Estos juicios van desde la primera persona, el Yo, el propio profesor que hace constantemente una autoevaluación de su práctica, sus decisiones y percepciones, hasta la emisión de juicios acerca de sus alumnos, colegas e instituciones.

El juicio al alumno es un recurso común en el discurso del profesor, como explicación o justificación de su propio actuar. Los juicios hacia los alumnos se dan acerca de su conducta, de sus habilidades o capacidades, o hacia los resultados obtenidos por éste en los exámenes o evaluaciones y que en ocasiones se manifiestan en discrepancias entre la valoración del profesor y la obtenida de la aplicación de un instrumento de evaluación.

Los juicios se conjugan con valoraciones realizadas por los profesores a diferentes aspectos de la práctica evaluativa y que se dan como positivas o negativas. Entre los aspectos que son objetos de valoración se encuentran los exámenes realizados por ellos mismos o por terceros, específicamente colegas, el proceso de diseño y los tipos de tareas de los exámenes. Entre los aspectos valorados de manera positiva destaca el de la colaboración, específicamente con colegas. Por otro lado, la práctica evaluativa de los profesores, en etapas tempranas de su carrera profesional, es percibida de manera negativa, reiteradamente por varios participantes.

Los profesores hacen valoraciones de otras prácticas, como obligatorias, justas o importantes. Entre los aspectos valorados como obligatorios se encuentran diseñar exámenes, evaluar habilidades, validez en los instrumentos, evidencias de aprendizaje, desempeñar prácticas de evaluación deseables, considerar el impacto de sus instrumentos y otras acciones relacionadas con las instituciones donde laboran. Los profesores valoraron como justas acciones como evaluar asegurando la validez de contenido, confiabilidad e igualdad, compensar al alumno por errores de parte suya, evitar abusar de su poder y asegurar que las calificaciones que obtengan sus alumnos sean justas. Otra valoración común fue la etiqueta de 'importante'. Entre los aspectos de la evaluación que consideran importantes se encuentran el mismo proceso de evaluación, la gramática y el vocabulario como componentes de la lengua y objetos de evaluación, el trabajo colaborativo, la preparación técnica, la validez de los instrumentos y la formación personal. Es así, que los profesores construyen este aspecto de su ética, a través de juicios y valoraciones, sobresaliendo los aspectos que son calificados como obligatorios, justos e importantes.

En cuanto a los juicios, cabe señalar que incluso aquellos que podrían considerarse como negativos, representan parte de la construcción de esta dimensión ética, visto desde el punto de vista de la formación del docente, de la experiencia y de las decisiones que en su propia

práctica deriven estos juicios. Es decir, una valoración negativa acerca de una práctica de un tercero, como un colega, bien podría ser considerada poco ética, sin embargo, si esta crítica conlleva una reflexión que guíe hacia una mejora en la propia práctica el desarrollo del profesional prevalece.

Las dos subdimensiones de la dimensión ética, la interpersonal y la valorativa se entrelazan entre sí y con otras dimensiones de la responsabilidad profesional, como la técnica y la lingüística. Entre sí, las dos subdimensiones comparten el aspecto de alteridad, el solo juicio del alumno o de las prácticas de los colegas tienen en sí mismas ese reconocimiento a la existencia del Otro. Nueva, e ineludiblemente, los aspectos técnicos de validez y confiabilidad son relacionados con el actuar del profesor, del deber, de la obligación y la importancia; así como las prácticas evaluativas, tanto propias como ajenas. El aspecto lingüístico se ve presente, aunque en menor grado, en la importancia que representan la gramática y el vocabulario en las prácticas docente y evaluativa.

Cabe mencionar algunos aspectos que surgieron en las respuestas que son ajenos al aprendizaje de lengua, como estrategias de resolución de tareas y aspectos o rasgos personales de los alumnos, como motivación o interés. Esto es evidencia de que la visión de los profesores no se limita al aprendizaje de la lengua, sino de una percepción mucho más amplia, de la condición humana del alumno que va más allá de los aspectos técnicos y lingüísticos. Esto al aplicarse a la práctica docente suma al aspecto ético y repercute en una práctica más profesional. Por otro lado, también existieron los juicios, predisposiciones y generalizaciones de parte de los profesores, que de manera paradójica serían en detrimento de una práctica ética y profesional.

5.4.2 La alteridad

En el análisis cualitativo se encontró que el segundo código con mayor presencia en las citas fue la alteridad y se ubicó dentro de la dimensión ética. Se refiere al reconocimiento o referencia de otro, mediante un posicionamiento en su lugar.

A la luz de los resultados, la alteridad se ha convertido en un factor esencial en la estructuración de la cognición del profesor sobre el aspecto evaluativo. Muchas de las decisiones que toma el profesor se relacionan con el Otro, generalmente el alumno y en

ocasiones un colega, consideración que suma al aspecto ético de la práctica evaluativa. La alteridad es una manera en que el profesor desarrolla identidad y comunidad (Ibarra, 2013), reconoce que su trabajo está ligado con el Otro, implica una separación del Yo y repercuten en un bienestar para el Nosotros.

La práctica evaluativa se enriquece cuando el profesor se ubica en el lugar del alumno, advierte cierta conducta o acciones por parte de este; o a través de su experiencia prevé cómo se siente o puede sentirse en una situación dada. La alteridad contribuye en la explicación de las razones para las acciones del profesor, cómo trata de dar las de sus alumnos. El profesor actúa con base en el Otro, por ejemplo en la búsqueda del logro de las cualidades y utilidad de un instrumento, como plantean Bachman y Palmer (1996), las acciones repercuten en un beneficio común, tanto para el profesor que desarrolla instrumentos válidos y confiables, como para el alumno que recibe el resultado de una evaluación justa.

Es decir, los conceptos de validez de constructo y confiabilidad en su aplicación práctica generan alteridad. Al definir el constructo a medir y considerar el impacto que tendrá dicha medición, además de considerar todos los aspectos técnicos que permiten obtener un resultado confiable y justo en beneficio del Otro, en este caso nuevamente, el alumno, necesariamente se origina alteridad, que en su manifestación equilibrada o próxima es ética (González Silva, 2009).

No obstante, la alteridad puede tener dos lados opuestos en su manifestación, como señala González Silva (2009), de manera que la alteridad equilibrada o próxima contribuye al aspecto ético de la práctica, sin embargo, la alteridad alterada o desvinculante tiene el efecto contrario. El discurso de los profesores en este caso muestra ambos tipos, que son parte de la dimensión ética; el primero que se muestra con el código Alteridad, el segundo podría relacionarse más con el código Juicio al alumno, que, si bien es un reconocimiento de la existencia y el papel del alumno, también es una valoración realizada desde solo una perspectiva y las limitaciones que puede tener el profesor del alumno.

En este estudio se ha identificado a la alteridad como un factor de decisión dentro de la dimensión ética, se ha advertido su presencia en el discurso de los profesores, no obstante, la exploración de la injerencia o el impacto de la alteridad en el trabajo del profesor en el

momento de realizar evaluación del aprendizaje de lengua por parte de sus alumnos requiere de estudios con mayor profundidad.

5.5 Las prácticas evaluativas de los profesores de la Universidad.

En este apartado se explican, a partir de los resultados de la investigación, algunas de las prácticas, creencias y pensamientos de los profesores de lengua sobre la evaluación del aprendizaje, lo que responde a la pregunta sobre cuáles son las prácticas del profesor de lengua como diseñador de exámenes en el proceso de evaluación del aprendizaje. Para la descripción se han considerado las categorías obtenidas del análisis cualitativo, así como las dimensiones y el análisis temático. Al igual que en este último, los incisos representan los temas propuestos por el investigador a través de las preguntas realizadas en las entrevistas. De igual manera se discutirán estas prácticas a la luz de la teoría presentada sobre evaluación del aprendizaje de lenguas y diseño de exámenes de lengua.

Es preciso reiterar que la visión del proceso de evaluación del aprendizaje de lengua es de un sistema complejo, con las características señaladas previamente en este documento. La generalización y el establecimiento de patrones estadísticos no representan esta visión; aunque sí decir que los patrones encontrados, así como las coincidencias y excepciones obedecen a esas cualidades, opuestas y complementarias a la vez, que son la dependencia y la independencia. Es decir, factores en común como la formación profesional, el contexto, las políticas institucionales y las reglas de operación pueden dar lugar a las coincidencias en prácticas, pensamientos y creencias; pero la independencia de cada unidad, que es cada profesor, también da lugar a esas excepciones, así como a contradicciones propias de un sistema complejo.

5.5.1 El proceso de diseño de exámenes

Los profesores inician el proceso de diseño de exámenes, con una revisión de los contenidos y los temas estudiados durante el curso, lo que asegura validez de contenido en sus instrumentos. Con base en las habilidades a evaluar asignan o diseñan tareas y buscan un balance de ellas en la evaluación. Por balance, iniciativa propia o por alguna política institucional o departamental, algunos profesores y profesoras ya tienen una estructura y tareas predeterminadas. Otros aplican tareas similares a las que se utilizan para evaluar las

habilidades en los exámenes de certificación o se ayudan de las tareas de los exámenes incluidos en el libro de texto u otros libros de texto correspondientes al nivel.

Las prácticas evaluativas encontradas son congruentes con mayoría de la teoría señalada en la literatura acerca de proceso de diseño de exámenes y evaluación (Genesee y Upshur, 1996; Brown y Abeywickrama, 2018). No obstante, hay aspectos señalados en la teoría que escapan a la práctica de los participantes en el estudio. Por ejemplo, Brown y Abeywickrama (2018) señalan la necesidad de redactar las especificaciones de un examen, en la práctica se encontró que, si bien, solo uno de los profesores señaló prepararlas (Lorena), el resto realizan prácticas aisladas que hacen a la vez de especificaciones, como tener un formato prediseñado al que se adaptan las tareas.

Por otro lado, también se encontró que entre los profesores hay prácticas que no se encuentran en la teoría, que no se sugieren debido a que pudiesen reducir el grado de validez. Los resultados revelaron que entre los profesores hay quienes no diseñan o han diseñado muchos instrumentos de evaluación y en su lugar utilizan tareas diseñadas para exámenes de certificación o algunos incluidos en los libros de texto, principalmente entre los profesores de menor experiencia,. Esto podría representar un distanciamiento entre los objetivos instruccionales, la práctica diaria y evaluación formativa de la evaluación sumativa, aunque habría que realizar un estudio para comprobarlo.

Otro aspecto a resaltar en la práctica evaluativa fue que a pesar del enfoque de este estudio en el aspecto sumativo de la evaluación y en los exámenes en particular, los profesores abordaron el tema de la evaluación formativa en distintos momentos de las entrevistas, lo cual indica una consideración importante a este tipo de evaluación más cercano a la evaluación para el aprendizaje (Davison y Leung, 2009). Prácticas como la retroalimentación en busca del aprendizaje del alumno (Brown y Abeywickrama, 2018) se evidencian en las entrevistas, así como una variedad de tipos de evaluación. El arraigo de la evaluación sumativa es grande, debido a políticas institucionales e incluso en algunos entrevistados, signo del carácter burocrático de la evaluación (Rea-Dickins, 2001) sin embargo la práctica evaluativa para el aprendizaje y de cuidado del alumno es una parte importante en la práctica de los participantes, resaltando las identidades pedagógicas y de aprendizaje de la evaluación.

5.5.2 Características de un examen de lengua

En cuanto a las características que debe tener un examen de lengua los profesores mencionaron las siguientes:

- I. Cualidades del examen
 - Validez
 - Confiabilidad
 - Práctico y fácil de calificar
- II. Características relacionadas con la validez de constructo
 - Evaluar los contenidos o temas del curso o validez de contenido
 - Evaluar todas las habilidades
 - Contextualización
 - Validez visual
- III. Características relacionadas con la confiabilidad del instrumento
 - Variedad y equilibrio o balance entre las tareas
 - Equilibrio o balance en el grado de dificultad
 - El examen debe representar un reto para el alumno
 - Instrucciones claras

Lo anterior indica que a pesar de que algunos de los profesores entrevistados no manejaban la terminología técnica de diseño de exámenes y que para algunos otros no es clara, las características que surgieron del análisis sugieren que los profesores buscan en su práctica el logro de validez y confiabilidad. Estos resultados son congruentes con lo encontrado en la fase cuantitativa donde pudo notarse el énfasis en el logro de aquellos aspectos relacionados con la confiabilidad, principalmente, y de la validez. Más sobre el logro de estas dos cualidades de los instrumentos de evaluación en el siguiente inciso.

5.5.3 Las prácticas evaluativas y el logro de validez y confiabilidad

Parece no existir un consenso entre los profesores sobre lo que miden sus exámenes de lengua. Las respuestas en este sentido fueron diversas: se mide nivel de lengua, cumplimiento de objetivos, conocimiento, comprensión, dominio y competencia. El aprendizaje de lengua no fue mencionado, aunque bien podría explicarse utilizando este término como uno que englobe todos los mencionados por los profesores.

Con la finalidad de aumentar el grado de validez de los instrumentos de evaluación es necesario clarificar y consensar su objetivo, determinar con precisión qué significa cada uno de los términos surgidos en las respuestas y cómo se relacionan con el aprendizaje de lengua y con la competencia comunicativa. La mayoría de los problemas que los profesores encontraron en los exámenes y que generan dudas, se relacionan con la confiabilidad de los resultados, por lo que se recomienda que se realice un exhaustivo trabajo en cuanto a aumentar este aspecto de los instrumentos se refiere. Los resultados muestran que una revisión completa por parte de los profesores a los exámenes antes de aplicarlos no es una práctica que se realice en todos los casos, lo que ha contribuido a problemas de validez y confiabilidad en los instrumentos, lo que genera reacciones entre los profesores, como se explica a continuación.

5.5.4 Percepciones de los exámenes como instrumentos de evaluación

Entre las ideas que manifestaron los profesores con respecto a los exámenes estuvieron escepticismo, obligación, duda e incluso, pena y pereza, y por otro lado importancia, necesidad, gusto y admiración. De igual manera, como aspecto positivo, se pudo notar en el análisis una constante relación entre los exámenes y el cuidado del alumno por parte del profesor.

Se percibió entre los profesores la inquietud de señalar que los exámenes no son, o no deben ser la única forma de evaluar el aprendizaje de lengua. Las razones para esta afirmación acerca de la forma de evaluar y del uso de los exámenes como instrumentos fueron:

- Que en ocasiones no eran evaluaciones totalmente objetivas, o que no mostraban toda la información necesaria.

- Factores externos como el estrés, que afectan el desempeño del alumno en el examen.
- La dificultad que representa para el profesor que un examen incluya todos los contenidos del curso o una muestra importante y representativa de ellos.
- Errores de diseño que incluyen un amplio espectro de acciones en distintas etapas del diseño como el visual, instrucciones, tareas, reactivos.

Es de resaltar la duda que generan los exámenes entre algunos profesores, en el sentido que, como ellos lo expresan, no reflejan lo que deberían mostrar, ya sea conocimiento, nivel de lengua, competencia o desempeño en el salón de clase. Esta situación implica una falta de validez y confiabilidad en los instrumentos, que a su vez representa una ruptura importante en la esencia y función del instrumento. Las dudas sobre el uso de exámenes han sido causadas por diversas razones, como errores de diseño, falta de revisión del instrumento antes de su aplicación, instrucciones inadecuadas, tareas o reactivos con grado de dificultad superior al nivel de competencia de los alumnos.

En cuanto al constructo o al objeto de medición de los exámenes de lengua, parece no existir un consenso. Los profesores mencionaron que los exámenes sirven para conocer el nivel de lengua, cumplimiento de objetivos, conocimiento, comprensión, dominio, competencia, aprendizaje, avance, áreas de oportunidad, fortalezas, e incluso aspectos ajenos a la lengua como personalidad del alumno, motivación e interés. Al tratarse de clases de lengua, donde desarrollan y aplican evaluaciones, se podría pensar que lo que se busca medir es el aprendizaje, que el alumno haya cumplido un objetivo y desarrollado una competencia a lo largo de un periodo de tiempo durante el curso y que le permita tener una experiencia comunicativa exitosa. Clarificar el objeto de medición, de manera consensual pudiese contribuir al logro de evaluaciones con mayor grado de validez, y por ende más justas.

Por otro lado, se muestra en los resultados que los profesores notan una mejora en su diseño de exámenes de lengua comparados con los instrumentos elaborados en los inicios de sus carreras. Con la experiencia los profesores perciben sus exámenes como más justos, más balanceados, más válidos, más completos o, simplemente, mejores.

5.6 Impacto del estudio en la literacidad de evaluación de lenguas

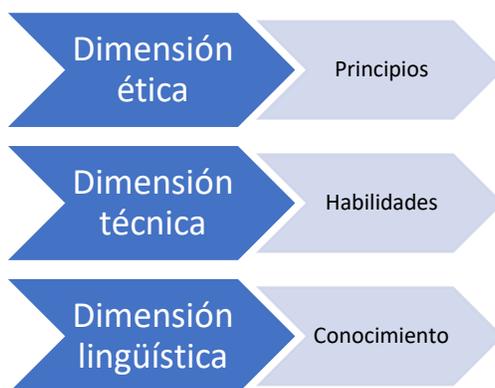
El objetivo de esta investigación no se enfocaba al estudio de la literacidad en evaluación de lenguas, sin embargo, dada la convergencia de aspectos de la LEL y dimensiones y factores de la RP, se considera pertinente profundizar en posibles repercusiones e impacto en estudios de literacidad de la evaluación de lenguas, e igualmente con ello responder a la pregunta sobre cómo se manifiesta esta literacidad en el discurso de los profesores.

En primer lugar, es necesario que el proceso de especialización en evaluación descrito en el inciso anterior se acelere, de manera que los profesores con menor experiencia, que son quienes muestran un menor grado de tecnicidad y conciencia de lo que implica la evaluación, lo desarrollen en etapas más tempranas de su carrera profesional.

Las decisiones del profesor tienen base en los aspectos técnicos, lingüísticos y principios éticos, que son similares o correspondientes a las tres dimensiones de la responsabilidad profesional, así como su literacidad en evaluación de lenguas (Davies, 2008, Taylor, 2009).

Figura 5.3

Dimensiones de la responsabilidad profesional y componentes de la literacidad en evaluación de lenguas



Nota: Elaboración propia.

Con base en los resultados obtenidos, es posible establecer que estos aspectos son resultado de un acumulado de experiencias del profesor, que no se limitan a su formación como tal,

sino que se remontan a los momentos iniciales de su aprendizaje de lengua. Cuando el profesor inicia el ejercicio de su práctica profesional, su primer referente es su experiencia como estudiante de lengua, a pesar de su formación como docente. La preparación en cuanto a aspectos técnicos de diseño de instrumentos de evaluación es insuficiente, y es en estas condiciones como el profesor realiza las evaluaciones. Ya durante el ejercicio docente, el profesor especializa su práctica con base en cursos, obtención de grados académicos, conocimiento empírico, trabajo colegiado y políticas y prácticas institucionales. Por ejemplo, algunos de los profesores entrevistados en este estudio indicaron que sus instrumentos se habían vuelto mejores con la experiencia.

Este estudio confirma varios de los hallazgos en estudios de literacidad en evaluación. Por ejemplo, se confirmó el enfoque ético y humanístico por parte de los evaluadores previsto como tendencia por Bachman (2000). Se confirma que los profesores requieren de una mayor preparación y conocimiento de la competencia comunicativa. Se coincide con Fulcher (2012) quien mencionó que los principios éticos guían la práctica y que los contextos nos ayudan a entenderla, así como algunas de las razones. Giraldo (2017) propone que entre las habilidades que se requieren del profesor, se encuentran la instruccional, la habilidad en diseño de evaluaciones y habilidad tecnológica, las cuales se hacen notar en diferentes entrevistas en este estudio. Por otro lado, la hipótesis de Hill (2017) sobre la carencia de habilidad del profesor para reflexionar sobre su práctica puede cuestionarse en este caso de estudio específico. El discurso de los profesores muestra una sustanciosa reflexión sobre su propio hacer y, a pesar de que es provocado por el mismo investigador, en el discurso hay evidencia de reflexión en momentos previos a la entrevista o durante la práctica. El ejercicio de provocar reflexión sobre la práctica demostró ser una herramienta que puede, y debe, utilizarse para identificar necesidades de preparación y desarrollo en evaluación y que contribuye a solventar esa carencia encontrada por Hill (2017).

La experiencia del profesor contribuye al logro de un mayor grado de validez y confiabilidad en el diseño de instrumentos, lo cual repercute en una práctica más ética, de religación, de responsabilidad y cuidado. El peso de esta responsabilidad profesional en las decisiones tomadas por el profesor está aún por determinarse, sin embargo, en este trabajo se muestra en el análisis del discurso de los profesores que la responsabilidad profesional es

una gran estructura que conlleva factores de diversa índole, que conjugados y categorizados conforman una estructura paralela a la literacidad en evaluación del profesor.

5.7 El carácter social de la evaluación en el salón de clase

El estudio muestra que la evaluación en el salón de clase no es ajena al carácter social de la evaluación a gran escala (McNamara, 2000; McNamara y Roeber, 2006). Se encontraron evidencias de la búsqueda de equidad (Kunnan, 2001), justicia, alteridad y cuidado del alumno. Estas cualidades, junto con todas las que integran la dimensión ética y la dimensión personal de la responsabilidad profesional dan cuenta del aspecto social.

El actuar del profesor en búsqueda del bienestar del alumno es un asomo del principio de religación de Morin (2006) y algunos principios de la evaluación moral de House (1997) también surgen al analizar las acciones del profesor, especialmente aquellas donde se da un diálogo y consideración de parte del profesor a los alumnos. Esto aunado a la aceptación, acogimiento y enfrentamiento de la alteridad de Lévinas (Quiroz y Marín, 2013), las cuales se encuentran de igual manera en la actitud del profesor, son todos promotores del carácter social de la evaluación del salón de clase.

El estudio cuantitativo evidenció la importancia tanto del aspecto ético como de la consciencia que tienen los profesores del impacto de sus instrumentos, sin embargo un acercamiento a la construcción de un argumento para la validez de los instrumentos, como proponen Bachman y Palmer (2010) y Bachman y Damböck (2018) pareciera ser una empresa lejana aún y esto resta al desarrollo del profesional, porque a pesar de ser un aspecto técnico se relaciona estrechamente con el bienestar del alumno; incluso un consenso en cuanto al objeto de evaluación por parte de los profesores es necesario, como ya se comentó previamente en este documento.

Así que no se pretende elaborar una apología del profesor, o una crítica por acciones poco deseables que son en detrimento de lo social, la religación o del bienestar ajeno, actos egoístas más que altruistas. Se percibe al profesor desde una mirada de paridad, como seres humanos falibles, dadas sus circunstancias, contextos o situaciones particulares, si bien sí con principios que guían hacia una ética, pero con una falibilidad característica de la condición humana.

El salón de clase es un contexto particular, donde el encuentro profesor-alumno se vuelve de carácter social, de construcción y de aprendizaje. El proceso de evaluación puede ser causal de desencuentro entre estos agentes, pero es el profesor en su posición de poder, el responsable de que no sea así, o de minimizar el efecto. Por otro lado, la práctica evaluativa implica una responsabilidad ante el gremio (Fulcher, 2013), ante la comunidad de profesores que el profesor a cargo del grupo de alumnos representa. De este modo, el carácter social, no es solo vertical, sino horizontal en una relación de colaboración y responsabilidad ante los colegas.

5.8 Cambio en el paradigma educativo sobre la evaluación de lenguas

El docente de lenguas es generalmente el principal responsable de recopilar la información para la toma de decisiones en el proceso de evaluación, y en el ámbito de la enseñanza de lenguas no es distinto, como lo comentan Genesee y Upshur (1996:3):

Más que nadie, los profesores están involucrados en la evaluación de segundas lenguas, algunas veces como la persona a cargo de tomar decisiones, algunas otras recopilando la información necesaria para que otros las tomen, y otras más auxiliando a otros a tomarlas ofreciendo interpretaciones acerca del desempeño del alumno.¹¹

De aquí la importancia de que el profesor de lengua a cargo de la evaluación pueda proporcionar información suficiente para la toma de dichas decisiones y, como se menciona párrafos arriba, en la enseñanza de lenguas encuentra en la aplicación de exámenes su forma predominante de evaluación.

No obstante, el diseño de exámenes de lenguas requiere un alto grado de especialización (Bachman y Palmer, 2010:1) y a pesar de los avances logrados en materia de diseño de exámenes de lengua desde los años ochenta, aún existen vacíos y es necesaria más investigación para llenarlos, como reducir los cuestionamientos en cuanto a validez, especialmente (Bachman, 2000) y promover la evaluación para el aprendizaje. El trabajo del profesor de lenguas como responsable de la evaluación se enriquecería con una variedad de técnicas y estrategias de recopilación de información que se apeguen más al paradigma constructivista, que respondan a una necesidad de acercamiento de los

¹¹ Traducción del autor.

instrumentos de evaluación al uso real de la lengua (Bachman y Palmer, 2010) y como resultado, una evaluación para el aprendizaje de la lengua. De aquí la importancia puntualizada por Davies (2008) y Taylor (2009) de la literacidad en evaluación de lenguas, que es el camino hacia la evolución o desarrollo del papel del profesor, la profesionalización como evaluador

De esta manera, la profesionalización obliga a desarrollar el carácter técnico, donde la validez y la confiabilidad se amalgaman con el trabajo en el aula y en el día a día del profesor de lengua, sus valores, ética personal, profesional y experiencia. Como puede verse en los resultados y en el discurso de los profesores, estos aspectos son objeto de un sinnúmero de prácticas que, de manera consciente o empírica, llevan a una práctica más profesional, con mayor grado de especialización y de responsabilidad.

De esta manera, la labor del profesor de lenguas toma en una dimensión adicional un carácter más humanista e integral, en la que se conjugan las dimensiones descritas primero en la persona y enseguida en el profesional y especialista. Una dimensión humana en la que la práctica docente y la evaluativa se conforman con el quehacer diario, con la experiencia, la convivencia y comunidad.

CONCLUSIONES

En este apartado se dará cuenta de los que se consideran los principales hallazgos y aportaciones de este trabajo de investigación. De igual manera, de sus limitaciones, posibles brechas de estudios que pudiesen derivar y aquellos que se consideran necesarios para clarificar o consolidar la teoría propuesta.

Entre los resultados que se consideran más relevantes se encuentra la conformación y propuesta de la *responsabilidad profesional* como un sistema complejo para la toma de decisiones por parte del profesor en la práctica evaluativa del aprendizaje de lenguas. El discurso de los profesores ha dado evidencia de cómo los profesores estructuran su práctica evaluativa con base en tres dimensiones principales, la ética, técnica y lingüística. Se etiquetaron códigos que describen y estructuran cada una de las dimensiones y cabe resaltar la estructura de la dimensión ética en dos subdimensiones, que a su vez describen cómo se conforma la ética del profesional de la docencia de lenguas: La subdimensión interpersonal y la valorativa.

De esta manera se pone sobre la mesa el concepto *responsabilidad profesional* la cual debe aún explorarse, de manera que dé cuenta de cómo funciona en otros casos y contextos, por ejemplo, en diferentes niveles educativos, en escuelas del sector privado, incluyendo centros de lenguas. Es necesario desarrollar instrumentos que permitan conocer lo anterior y obtener resultados de manera práctica para el investigador y que coadyuven a la toma de decisiones por parte de los directivos, coordinadores y personas involucradas en la conducción de los procesos evaluativos.

Gracias a este estudio fue posible notar el enriquecimiento de la cognición de los profesores con la reflexión sobre el proceso y sus propias prácticas. La misma *responsabilidad profesional*, que como sistema complejo se autorregula, redirige, autoalimenta, se desarrolla con la reflexión de los docentes. Se recomienda el uso de entrevistas y protocolos verbales para el estudio de procesos de evaluación, ya que permitirían una mirada mucho más introspectiva, ampliar la documentación sobre las prácticas reales en diferentes contextos y fomentar la práctica reflexiva.

La aplicación de distintos enfoques metodológicos a los aplicados en este estudio de caso enriquecería los resultados obtenidos. Estudios longitudinales en lugar de transversales, así como estudios etnográficos, o de investigación-acción podrían dar cuenta de manera más amplia acerca de la práctica evaluativa y la responsabilidad profesional.

La perspectiva del proceso evaluativo como un sistema complejo podría ayudar a entender la práctica y mecanismos de desarrollo profesional y especialización del profesor de lengua. La literacidad en evaluación de lenguas (LEL) está vinculada estrechamente con la *responsabilidad profesional* (RP), el desarrollo de la primera tendrá implicaciones en la segunda y viceversa. Se requiere investigación acerca de las conexiones entre estos dos sistemas para generar un desarrollo profesional más eficiente y cercano a los profesores, al identificar casos, necesidades y prácticas particulares. Otra arista de la visión de la práctica evaluativa como un sistema complejo llevaría necesariamente a replantear procesos de enseñanza de lengua, en congruencia también con la visión de las lenguas como sistemas complejos, como ha planteado Larsen-Freeman (2011) y la coherencia que idealmente debe tener la evaluación con la metodología y los objetivos de los cursos de lenguas planteados desde la propia institución.

El estudio ha revelado el alto grado de consideración de los aspectos éticos en el discurso y las prácticas de los profesores, así como una serie de reflexiones, descripciones de prácticas ideas y conceptos sobre la evaluación de lenguas. No obstante, el estudio se ha enfocado exclusivamente en el profesor de lengua, mostrando solamente uno de los puntos de vista, de tantas miradas que conlleva el proceso evaluativo. Se necesita investigación que corrobore los dichos y prácticas de los profesores entrevistados y las prácticas e ideas identificadas en la encuesta, a través de conocer la perspectiva de administradores, directivos y alumnos. La triangulación de la práctica de diseño de exámenes descrita por los profesores por medio de observación de clases, análisis de instrumentos y acompañamiento del docente en el proceso, también es necesaria para dar mayor validez a los resultados obtenidos.

El enfoque del estudio desde su planteamiento original fue en la evaluación sumativa, específicamente en el diseño de exámenes, no obstante, los profesores trajeron a colación el tema de otras formas e instrumentos de evaluación, lo que dio lugar a hablar de práctica

evaluativa. No obstante, más estudios se requieren para incorporar al sistema de evaluación, otros tipos de evaluación, específicamente formativa, además de continua y alternativa, aunque no de manera tangencial, como se dio en este caso de estudio.

El estudio tuvo en su fase cuantitativa la limitante del número de profesores encuestados. El reducido número de participantes en esta fase ($n=88$) aumentó el margen de error en la estadística, además de disminuir el índice KMO, lo que tuvo un efecto en la confiabilidad interna del instrumento en el momento de realizar el análisis factorial. A pesar de no ser congruente con la epistemología utilizada como base en este estudio, se requiere de un estudio cuantitativo sobre las prácticas evaluativas y de diseño de exámenes de los profesores de la universidad, con una muestra más amplia. El estudio también se limitó a la zona urbana de la ciudad de Toluca, cuando la universidad tiene planteles a lo largo de todo el Estado de México. Ampliar el estudio a más sujetos y a estos planteles permitiría mayor conocimiento de las prácticas evaluativas de los profesores de la universidad en general y podrían hacerse estudios comparativos entre diferentes áreas geográficas. De igual manera, averiguar cómo opera la *responsabilidad profesional* y cómo se comparan las prácticas evaluativas de la universidad en cuestión con otras universidades de otros lugares del país, así como dar luz sobre en qué grado se relaciona la evaluación con el contexto institucional y regional.

La evaluación del aprendizaje ha sido un aspecto generalmente relegado por los profesores, administradores, directivos de instituciones educativas y por los planeadores en el sistema educativo en general y de enseñanza de lenguas. Si bien, existe un consenso en cuanto a su importancia se refiere, poco se hace para el desarrollo del aspecto evaluativo en los profesores. Los participantes de la fase cualitativa de este estudio manifestaron haber dedicado poco tiempo al desarrollo de su literacidad en evaluación, lo que incluyó los planes curriculares en sus estudios de formación docente.

Un mayor énfasis debe darse en el desarrollo de la literacidad en evaluación y en la evaluación para el aprendizaje. Mientras la evaluación sumativa sea vista como la finalización de un proceso de enseñanza y no como un punto de partida o inicio de otro, seguirá existiendo una separación entre este tipo de evaluación y la formativa. La evaluación para el aprendizaje, o evaluación en el salón de clase, que desde la perspectiva

de Wang (2017) incluye ambos tipos de evaluación es necesaria. Se requiere del desarrollo de aspectos técnicos de diseño de exámenes, especialmente en la búsqueda de validez, la cual se logrará con ideas más claras y consensuadas sobre lo que debe evaluarse y si a la vez se logra en conjunto con la evaluación del aprendizaje de lengua, este sería mucho más representativo y exitoso, ya que daría lugar al desarrollo de un proceso más integral, donde la evaluación no está desligada de la enseñanza-aprendizaje, sino que son parte del mismo sistema, de manera mucho más cohesiva y en beneficio principalmente del alumno.

LISTA DE REFERENCIAS

- Alderson, J. C. (2000). *Assessing Reading*. Cambridge University Press.
- Alderson, J.C. y Banerjee, J. (2001). Language Testing and Assessment. *Language Teaching*, 34 (4), 213-236. [DOI: 10.1017/S0261444800014464](https://doi.org/10.1017/S0261444800014464)
- Alderson, J. C. y Hamp-Lyons, L. (1996). TOEFL Preparation Courses: A Study of Washback. [DOI: 10.1177/026553229601300304](https://doi.org/10.1177/026553229601300304)
- Alderson, J. C. y Wall, D. (1993). Does Washback Exist? *Applied Linguistics*. 14 (2), 115-129. [DOI:10.1093/applin/14.2.115](https://doi.org/10.1093/applin/14.2.115).
- Azuaje, L. y González, M. (2018). Reflexiones sobre la epistemología, axiología y ontología de la investigación docente. *CIEG, Revista Arbitrada del Centro de Investigación y Estudios Gerenciales*. (33), 251-259.
- Bachman, L. (2000). Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing* 2000, 17:1. [DOI: 10.1177/026553220001700101](https://doi.org/10.1177/026553220001700101)
- Bachman, L y Damböck, B. (2018). *Language Assessment for Classroom Teachers*. Oxford University Press.
- Bachman, L. y Palmer, A. (1996). *Language Testing in Practice*. Cambridge University Press.
- Bachman, L. y Palmer, A. (2010). *Language Assessment in Practice*. Cambridge University Press.
- Bechmann, G. y Stehr, N. (2002). The Legacy of Niklas Luhmann. *Society* 39: 69–75. <https://doi.org/10.1007/BF02717531>
- Bordón, T. (2015). La evaluación de segundas lenguas (L2). Balance y perspectivas. *Revista Internacional de Lenguas Extranjeras* (4), 9-30. <https://doi.org/10.17345/rile4.695>.
- Borón, A. (2006). Prólogo. En Sotolongo, P. y Delgado, C. (Eds.), *La revolución contemporánea del saber y la complejidad social: Hacia unas ciencias sociales de nuevo tipo*. Colección Campus Virtual.
- Brown, H.D. (2003). *Language Assessment: Principles and Classroom Practices*. Longman.
- Brown, J. (2014). *Mixed methods research for TESOL*. Edinburgh University Press.
- Brown, H.D. y Abeywickrama, P. (2018). *Language Assessment: Principles and Classroom Practices*. Longman.
- Buck, G. (2001). *Assessing Listening*. Cambridge University Press.

- Buenfil Burgos, R.N. (1993). Análisis de Discurso y Educación. Conferencia presentada en el Centro de Investigación Educativa de la Universidad de Guadalajara. DIE-CINVESTAV.
- Cabra-Torres F. (2010). El diálogo como fundamento de comunicación ética en la evaluación. *Educación-Educadores*. Bogotá, Colombia. 2010;13(2): 239-252. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-12942010000200005
- Cai, H. (2013). Partial dictation as a measure of EFL listening proficiency: Evidence from confirmatory factor analysis. *Language Testing*, 30(2), 177–199. <https://doi.org/10.1177/0265532212456833>
- Canale, M. y Swain, M. (1980). Theoretical bases of communicative approaches to Second language teaching and testing. *Applied Linguistics*. 1(1), 1–48.
- Canale, M. (1983). From communicative competence to communicative language pedagogy. En Richards J., y Schmidt R. (eds.). *Language and Communication*. Longman.
- Cappelletti, I. (2004). *Evaluación educativa: Fundamentos y prácticas*. Siglo XXI.
- Casero Martínez, A. (2010). ¿Cómo es el buen profesor universitario según el alumnado? *Revista española de pedagogía*. Año LXVIII, No. 246, mayo agosto 2010, 223-242
- Celce-Murcia, M. (2008). Rethinking the Role of Communicative Competence in Language Teaching. En Alcón Soler E. y Jordá M.P.S., *Intercultural Language Use and Language Learning*. Springer.
- Cushing Weigle, S. (2002). *Assessing Writing*. Cambridge University Press.
- Chacón-Arteaga, N. L. (2014). El enfoque ético, axiológico y humanista aplicado a la educación. *Varona*, 59, 14-22.
- Chalhoub-Deville. M. (2016). Validity Theory: Reform, Policies, Accountability Testing and Consequences. *Language Testing*. 33(4), 453-472.
- Chapelle, C. (2011). Validity Argument for Language Assessment: The Framework is Simple. *Language Testing*, 29(1), 19-27. [DOI: 10.1177/0265532211417211](https://doi.org/10.1177/0265532211417211)
- Chapelle, C. (2013). Conceptions of Validity. En *The Routledge Handbook of Language Testing*, Fulcher, G. y Davidson, F. (Eds.). Routledge. [doi/10.4324/9780203181287.ch3](https://doi.org/10.4324/9780203181287.ch3)
- Chapelle, C. y Douglas, D. (2006). *Assessing Language through Computer Technology*. Cambridge University Press.
- Cheng, L. (1997). How Does Washback Influence Teaching? Implications for Hong Kong. *Language and Education*, 11(1), 38-54. [doi/10.1080/09500789708666717](https://doi.org/10.1080/09500789708666717)
- Creswell, J.W. (1998). *Qualitative Inquiry and Research Design*. Sage Publications.

- Creswell, J.W. (2014). *Research Design: Qualitative, Quantitative; and Mixed Methods Approaches*. Sage Publications.
- Creswell, J.W. y Plano Clark, V.L. (2018). *Designing and Conducting Mixed Methods Research*. Sage Publications.
- Davidson, P. (2012). *The Cambridge Guide to Second Language Assessment*. Cambridge University Press.
- Davies, A. (2004). Language Testing and the Golden Rule. En *Language Assessment Quarterly*, 1(2 y 3), 97-107. <http://dx.doi.org/10.1080/15434303.2004.9671778>.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327–347.
- Davies, A. (2012). Kane, Validity and Soundness. *Language Testing*, 29(1), 37–42. [doi/10.1177/0265532211417213](http://dx.doi.org/10.1177/0265532211417213).
- Davies, A. (2013). Fifty Years of Language Assessment. En *The Companion to Language Assessment*. Kunnan, A. (Ed.). <https://doi.org/10.1002/9781118411360.wbcla127>.
- Davies, A. y Elder, C. (2005). Validity and Validation in Language Testing. En Hinkel, E. (Ed) *Handbook of Research in Second Language Teaching and Learning*. Routledge.
- Davison, C. y Leung, C. (2009). Current issues in English Language Teacher-Based Assessment. *TESOL Quarterly*, 43(3), 393-415. <https://doi.org/10.1002/j.1545-7249.2009.tb00242.x>
- De Almeida, (2008). *Para entender la complejidad*. Multiversidad Mundo Real Edgar Morin, A.C.
- Deygers, B., Van den Branden, K., y Van Gorp, K. (2017). University entrance language tests: A matter of justice. *Language Testing*. <https://doi.org/10.1177/0265532217706196>.
- Domínguez Prieto, X. M. (2003). *Ética del docente*. Editorial Mounier.
- Douglas, D. (2000). *Assessing Languages for Specific Purposes*. Cambridge University Press.
- Douglas, D. (2014). *Understanding Language Testing*. Routledge.
- Espinosa-Luna, C. (2013). La teoría de los sistemas de Niklas Luhmann a prueba. Horizontes de su aplicación en la investigación social en América Latina. *Sociológica* (México) 28(78). http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-01732013000100009&lng=es&tlng=es.
- Farhady, H. (2009). *Ethics in Language Testing*. <http://mpazhou.ir/wp-content/uploads/2011/11/Ethics-in-Language-Testing.pdf>

- Flick, U. (2011). *Introducing Research Methodology*. A Beginner's Guide to Doing a Research Methodology. London: SAGE Publications.
- Ferreya, H.A. y Rúa, A. (2018). Dimensiones involucradas en el estudio de las prácticas de enseñanza. *Educación y Humanismo*, 20(34), 140-155. <https://doi.org/10.17081/eduhum.20.34.2861>.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. DOI: [10.1080/15434303.2011.642041](https://doi.org/10.1080/15434303.2011.642041)
- Fulcher, G. (2013). *Practical Language Testing*. Routledge.
- Fulcher, D. & Davidson, F. (2007). *Language Testing and Assessment: An Advanced Resource Book*. Routledge.
- Genesee, F. y Upshur, J. (1996). *Classroom Based Evaluation in Second Language Education*. Cambridge University Press.
- Giraldo, F. (2017). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20 (1), 179-195. DOI: [10.15446/profile.v20n1.62089](https://doi.org/10.15446/profile.v20n1.62089).
- Glaser, B., y Strauss, A. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Sociology Press.
- González Maura, V. (2004). El profesorado universitario: su concepción y formación como modelo de actuación ética y profesional. *Revista Iberoamericana de Educación*, 34(1), 1-11. DOI: <https://doi.org/10.35362/rie3412934>
- González de Rivera, J.L. (2004). Empatía y eempatía. *Psiquis: Revista de psiquiatría, psicología médica y psicosomática*, 25(6), 5-7. <https://dialnet.unirioja.es/servlet/articulo?codigo=1069291>
- González Silva, F. (2009). Itinerario de alteridad. Una reconstrucción para nuevas aproximaciones Fermentum. *Revista Venezolana de Sociología y Antropología*, 56(19), 616-641. <https://www.redalyc.org/pdf/705/70517761011.pdf>
- Grabowski, K. y Dakin, J.W. (2013). Test Development Literacy. En *The Companion to Language Assessment*. Kunnan, A. (Ed.). DOI: [10.1002/9781118411360.wbcla129](https://doi.org/10.1002/9781118411360.wbcla129)
- Green, A. (2006). Washback to the learner: Learner and teacher perspectives on IELTS preparation course expectations and outcomes. *Assessing Writing*, 11 (2), 113-134. DOI: [10.1016/j.asw.2006.07.002](https://doi.org/10.1016/j.asw.2006.07.002)
- Green, A. (2013). *Exploring Language Assessment and Testing: Language in Action*. Abingdon, U.K.: Routledge.
- Hamp-Lyons, L. (2000). Social, Professional and Individual Responsibility in Language Testing. *System* 28, 579-591. DOI: [10.1016/S0346-251X\(00\)00039-7](https://doi.org/10.1016/S0346-251X(00)00039-7)
- Harmer, J. (2015). *The Practice of English Language Teaching*. Pearson Education.

- Hernández Sampieri, R., Fernández-Collado, C. y Baptista, P. (2006). *Metodología de la Investigación*. McGraw-Hill Interamericana.
- Hill, K. (2017). Understanding classroom-based assessment practices: A pre-condition for teacher assessment literacy. *Papers in Language Testing and Assessment*, 6(1). http://www.altanz.org/uploads/5/9/0/8/5908292/3.silhill_final_formatted_proofed.pdf
- Hill, K. y McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), <https://doi.org/10.1177/0265532211428317>
- House, E.R. (1997). *Evaluación, ética y poder*. Morata.
- Hughes, A. (2003). *Testing for Language Teachers*. Cambridge University Press.
- Ibarra, G. (2005). Ética y formación profesional integral. En *Reencuentro*, 43.
- Ibarra, G. (2013). Ética de la enseñanza. En *Reencuentro*, 34-41.
- Ibarra, G. (2018). Ética profesional y cívica para la formación universitaria. *Revista Boletín Redipe*, 4(7), 67-77. <https://revista.redipe.org/index.php/1/article/view/372>
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385–402.
- Inbar-Lourie, O. (2016). Language Assessment Literacy. En Shohamy, E., Iair, G. y May, S. (Eds). *Language Testing and Assessment. Encyclopedia of Language and Education*. Springer International Publishing. DOI: [10.1007/978-3-319-02326-7_19-1](https://doi.org/10.1007/978-3-319-02326-7_19-1)
- James, C. (1993). What is Applied Linguistics? *International Journal of Applied Linguistics*. 3(1) 17-32. DOI: [10.1111/j.1473-4192.1993.tb00041.x](https://doi.org/10.1111/j.1473-4192.1993.tb00041.x)
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing*, 30(3), 345-362. DOI: [10.1177/0265532213480334](https://doi.org/10.1177/0265532213480334)
- Johnston, B. (2003). *Values in English Language Teaching*. Lawrence Earlbaum Associates.
- Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527-535. doi.org/10.1037/0033-2909.112.3.527
- Kane, M. (2001). Current concerns in Validity Theory. *Journal of Educational Measurement*, 38(4), 319-342.
- Krumm, H.J. (2007). Profiles Instead of Levels: The CEFR and Its (Ab) Uses in the Context of Migration. *The Modern Language Journal*, 91(4). 667-669.
- Kunnan, A.J. (2000). Fairness and Justice for All. En Kunnan, A., *Fairness and Validation in Language Assessment* (Ed.). Cambridge: Cambridge University Press.

- Kunnan, A. (2001). Test Fairness. En *Studies in Language Testing: European Language Testing in a Global Context*, Milanovic, M. y Weir, C. (Eds.). Cambridge University Press.
- Kunnan, A.J., (2004). Test Fairness. En Milanovic, M. y Weir, C. (Eds.) *European Language Testing in a Global Context*. Cambridge University Press.
- Kunnan, A.J. (2013). Approaches to Validation in Language Assessment. En Kunnan, A.J. (Ed) *Validation in Language Assessment*. Routledge.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169–197. DOI: [10.1177/0265532214554321](https://doi.org/10.1177/0265532214554321).
- Larsen-Freeman, D. (2011). A Complexity Theory Approach to Second Language Development/Acquisition. En: Atkinson, D. (Ed.). *Alternative Approaches to Second Language Acquisition*. Taylor & Francis.
- Ledesma, R., Ferrando, P., y Tosi, J. (2019). Uso del análisis factorial exploratorio en RIDEP. Revisores para autores y revisores. *Revista Iberoamericana de Diagnóstico y Evaluación - e Avaliação Psicológica*, 3(52),173-180. <https://doi.org/10.21865/RIDEP52.3.13>
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., y Tomás-Marco, I. (2014). El Análisis Factorial Exploratorio de los Ítems: una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151-1169. <https://dx.doi.org/10.6018/analesps.30.3.199361>
- Luoma, S. (2004). *Assessing Speaking*. Cambridge University Press.
- Martínez Navarro, E. (2010). *Ética profesional de los profesores*. Editorial Desclée.
- Mateo, J. y Martínez, F. (2008). *Medición y evaluación educativa*. Editorial La Muralla.
- Maturana, H. y Varela, F.J. (1994). *De máquinas y seres vivos*. Editorial Universitaria.
- Mauri, M. (2003). Los aspectos morales de la docencia universitaria. En Román B. (2003) *Por una ética docente*. Grafite Ediciones.
- McCay, P. (2006). *Assessing Young Learners*. Cambridge University Press.
- McNamara, T. (2000). *Language Testing*. Oxford University Press.
- McNamara, T. (2001). Language Assessment as Social Practice: Challenges for research. *Language Testing*. 2001, 18(4) 333-349 DOI:[10.1177/026553220101800402](https://doi.org/10.1177/026553220101800402)
- McNamara, T. y Roever, C. (2006). *Language Testing: The Social Dimension*. Oxford: Blackwell Publishing Ltd.
- McNamara, T. y Shohamy, E. (2008). Language Tests and Human Rights. *International Journal of Applied Linguistics*, 18-1.

- Méndez Martínez, C. y Rondón Sepúlveda, M.A. (2012). Introducción al análisis factorial exploratorio. *Revista Colombiana de Psiquiatría*, 41(1),197-207.
[https://doi.org/10.1016/S0034-7450\(14\)60077-9](https://doi.org/10.1016/S0034-7450(14)60077-9)
- Mendizábal, N. (2006). *Los componentes del diseño flexible en la investigación cualitativa*. En Vasilachis, I. (Ed.), *Estrategias de investigación cualitativa*. Editorial Gedisa.
- Mendoza Ramos, A. (2015). La validez en los exámenes de alto impacto. Un enfoque desde la lógica argumentativa. *Perfiles Educativos*, XXXVII (149), 169-186
http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-26982015000300010&lng=es&tlng=es.
- Mertens, D. (2014). *Research and Evaluation in Education and Psychology: Integrating Diversity with Quantitative, Qualitative, and Mixed Methods*. Sage Publications.
- Messick, S. (1989). Validity. En Linn, R.L., editor, *Educational measurement*. New York: Macmillan.
- Messick, S. (1990). *Validity of Test Interpretation and Use*. Educational Testing Services.
- Messick, S. (1996). Validity and Washback in Language Testing. *Language Testing*, 13 (3), 241-256. DOI: [10.1177/026553229601300302](https://doi.org/10.1177/026553229601300302)
- Moore, P., Grondin, G. y Andrade, H. (2020). *Dimensiones humanistas en la construcción narrativa de la identidad del docente de segundas lenguas*. Ediciones y gráficos Eón, Universidad Autónoma del Estado de México.
<http://hdl.handle.net/20.500.11799/109020>
- Moreno Olivos, T. (2016). *Evaluación del Aprendizaje y para el aprendizaje. Reinventar la evaluación en el aula*. Universidad Autónoma Metropolitana.
- Morin, E. (1990). *Introducción al pensamiento complejo*. Editorial Gedisa.
- Morin, E. (2004). La epistemología de la complejidad. *Gazeta de Antropología*, 20. DOI: [10.30827/Digibug.7253](https://doi.org/10.30827/Digibug.7253)
- Morin, E. (2006). *El Método 6: Ética*. Ediciones Cátedra.
- Navarro Loli, J.S., y Merino Soto, C. A. (2010). Un réquiem para la regla de Kaiser (eigen > 1) en la retención del número de factores. *Fractal: Revista de Psicología*, 22(3), 640. <https://doi.org/10.1590/S1984-02922010000900014>
- Ortega Ruiz P. (2004), La educación moral como pedagogía de la alteridad. *Revista Española de Pedagogía*. Año LXII, No 227.
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363-380. DOI: [10.1177/0265532213480336](https://doi.org/10.1177/0265532213480336)
- Pereira, Z. (2011). Los Diseños del método mixto en la investigación en educación: Una experiencia concreta. *Revista Electrónica Educare XV*(1)15-29.
<https://doi.org/10.15359/ree.15-1.2>

- Pill, J. y Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 363–380. DOI: [10.1177/0265532213480337](https://doi.org/10.1177/0265532213480337)
- Pimienta, J. (2008). *Evaluación de los aprendizajes*. Pearson Educación.
- Popham, W. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48(1), 4-11. DOI: [10.1080/00405840802577536](https://doi.org/10.1080/00405840802577536)
- Prigogine, I. (1995). Ciencia y azar. *Zona Erógena*. 23, 1-6.
[https://aprendizaje.mec.edu.py/aprendizaje/system/content/0c59c97/content/Prigogine,%20Ilya%20\(1917-2003\)/Prigogine,%20Ilya%20-%20Ciencia%20y%20azar.pdf](https://aprendizaje.mec.edu.py/aprendizaje/system/content/0c59c97/content/Prigogine,%20Ilya%20(1917-2003)/Prigogine,%20Ilya%20-%20Ciencia%20y%20azar.pdf)
- Purpura, J. (2004). *Assessing Grammar*. Cambridge University Press.
- Quiroz Lizarazo, E.H. y Marín López, A. (2013). Aportes de la ética de Emmanuel Lévinas para el concepto de responsabilidad social empresarial. *Revista Virtual Universidad Católica del Norte*, 38, 16-26.
<https://www.redalyc.org/articulo.oa?id=1942/194225730003>
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429-462.
<https://doi.org/10.1177/026553220101800407>
- Read, J. (2000). *Assessing Vocabulary*. Cambridge University Press.
- Riofrío, W. (2001). ¿Complejidad o simplicidad? En busca de la unidad de la ciencia. *A Parte Rei: revista de filosofía*, 16(9)
<https://dialnet.unirioja.es/servlet/articulo?codigo=3970767>
- Román, B. (2003). Apuntes para una ética del profesor universitario. En Román B. (2003) *Por una ética docente*. Grafite Ediciones.
- Romero Pérez C. (2003). Paradigma de la complejidad, modelos científicos y conocimiento educativo. *Agora digital*, 6. <https://dialnet.unirioja.es/servlet/articulo?codigo=925254>
- Sawaki, Y., Stricker, L. J., y Oranje, A. H. (2009). Factor structure of the TOEFL Internet-based test. *Language Testing*, 26(1), 005–030.
<https://doi.org/10.1177/0265532208097335>
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3) 309–327. DOI: [10.1177/0265532213480128](https://doi.org/10.1177/0265532213480128).
- Schwandt, T. y Jang, E. (2004). Linking Validity and Ethics in Language Testing: Insights from the Hermeneutic Turn in Social Science. *Studies in Educational Evaluation*, 30, 265-280.
- Shohamy, E. (1997). Testing methods, testing consequences: are they ethical? Are they fair? *Language Testing*, 14(3), 340-349. doi: [10.1177/026553229701400310](https://doi.org/10.1177/026553229701400310).

- Shohamy, E. (2000). The relationship between language testing and second language acquisition, revisited. *System*, 28(4), 541-553. [doi.org/10.1016/S0346-251X\(00\)00037-3](https://doi.org/10.1016/S0346-251X(00)00037-3)
- Shohamy, E. (2001a). The Social Responsibility of Language Testers. En Spolsky, B. et al. (Eds.), *New Perspectives and Issues in Educational Language Policy*. John Benjamin's Publishing.
- Shohamy, E. (2001b). *The power of tests: A critical perspective on the uses of language tests*. Longman.
- Shohamy E. (2007). The Power of Language Tests, The Power of the English Language and the Role of ELT. En: Cummins J., Davison C. (eds.) *International Handbook of English Language Teaching*. Springer International Handbooks of Education, vol 15, Boston: Springer.
- Skehan, P. (1988). State of the art: Language testing, part I. *Language Teaching*, 21(4), 211–21. [DOI: 10.1017/S0261444800005218](https://doi.org/10.1017/S0261444800005218)
- Skehan, P. (1989). State of the art: Language testing, part II. *Language Teaching*, 22 (1), 1–13. [DOI: 10.1017/S0261444800005346](https://doi.org/10.1017/S0261444800005346)
- Spolsky, B. (2008). Language Testing at 25: Maturity and Responsibility? *Language Testing* 25 (3), 297-305. [DOI: 10.1177/0265532208090153](https://doi.org/10.1177/0265532208090153)
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(3), 534–539.
- Stiggins, R.J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238-245.
- Strauss, A. y Corbin, J. (1998). *Basics of Qualitative Research Techniques*. Blackwell.
- Tagliante, C. (2005). *L'Evaluation*. Paris: CLE International.
- Tan, M y Turner, C.E. (2015). The Impact of Communication and Collaboration Between Test Developers and Teachers on a High-Stakes ESL Exam: Aligning External Assessment and Classroom Practices. *Language Assessment Quarterly*, 12(1), 29-49. <https://doi.org/10.1080/15434303.2014.1003301>
- Tashakkori, A. y Creswell J. (2007). The New Era of Mixed Methods Research. *Journal of Mixed Methods Research* 1(1), 3-7. [DOI 10.1177/2345678906293042](https://doi.org/10.1177/2345678906293042)
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154-155. [DOI: 10.1093/eltj/cci030](https://doi.org/10.1093/eltj/cci030)
- Taylor, L. (2009). Developing Assessment Literacy. *Annual Review of Applied Linguistics*, 29, 21-36. [DOI: 10.1017/S0267190509090035](https://doi.org/10.1017/S0267190509090035)
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some Reflections. *Language Testing*, 30(3), 403-412.

- Turner, C. E. (2017). Classroom Assessment. En Fulcher, G. y Davidson F. (Eds.) *The Routledge Handbook of Language Testing*. Routledge.
- Ur, P. (2012). *A course in English Language Teaching*. Cambridge University Press.
- UAEMex. (1984). *Reglamento de Facultades y Escuelas Profesionales de la Universidad Autónoma del Estado de México*. Universidad Autónoma del Estado de México. https://controlescolar.uaemex.mx/dce/Manuales/Legislacion_Universitaria/UAEM-ReglamentoDeFacultadesyEscuelasProfesionales.pdf
- UAEMex. (2022). *Agenda Estadística*. Universidad Autónoma del Estado de México. <http://planeacion.uaemex.mx/docs/AE/2021/AE2021.pdf>
- Vega, K. (2015). Construcción de la alteridad y el significado de la familia en las ventas ambulantes. *Cuadernos de Antropología*, 25(1), 25-46. DOI: [10.15517/cat.v25i1.18796](https://doi.org/10.15517/cat.v25i1.18796)
- Velasco Ariza, V., Anguiano Paliza, M. y Larrazolo Reyna, N. (2007). Propuesta metodológica para la formulación de una conceptualización del constructo de un examen de certificación del inglés como lengua extranjera. *REDIE. Revista Electrónica de Investigación Educativa*, 9(2), 0. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1607-40412007000200008&lng=es&tlng=es.
- Vogt, K. y Tzagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374–402. DOI/10.1080/15434303.2014.960046
- Vogt, K. Tzagari, D. y Spanoudis, G. (2020) What Do Teachers Think They Want? A Comparative Study of In-Service Language Teachers' Beliefs on LAL Training Needs, *Language Assessment Quarterly*, DOI: [10.1080/15434303.2020.1781128](https://doi.org/10.1080/15434303.2020.1781128)
- Wang, X. (2017). A Chinese EFL Teacher's Classroom Assessment Practices. *Language Assessment Quarterly*, 14(4), 312-327. <https://doi.org/10.1080/15434303.2017.1393819>
- Weir C.J. (2005) The Nature of Test Validity. En: *Language Testing and Validation. Research and Practice in Applied Linguistics*. Palgrave Macmillan, London. DOI: [10.1057/9780230514577_3](https://doi.org/10.1057/9780230514577_3)
- Yan, X., Cheng, L. y Ginther, A. (2018). Factor analysis for fairness: Examining the impact of task type and examinee L1 background on scores of an ITA speaking test. *Language Testing*, 36(2), 207–234. <https://doi.org/10.1177/0265532218775764>

ANEXOS

ANEXO A: REPORTE DE CÓDIGOS

https://drive.google.com/file/d/1GEc39XiUNrWwM-GLV8n0P-sxnHtw_buI/view?usp=sharing

ANEXO B: TABLA DE CO-OCURRENCIAS

https://docs.google.com/spreadsheets/d/1J_X4lZJNP-okx19ciLNwh91u-RhcaDId/edit?usp=sharing&oid=100406262656415828887&rtpof=true&sd=true

ANEXO C: MATRIZ CÓDIGO-PARTICIPANTE

<https://docs.google.com/spreadsheets/d/1zgh5R4cMDRSIrr7iA6myiAJ3h72fO1ZY/edit?usp=sharing&oid=100406262656415828887&rtpof=true&sd=true>

ANEXO D: REPORTE DE CÓDIGOS POR DIMENSIÓN

https://drive.google.com/file/d/1TO-axguhTAkfHzgqHrH_mRfw3bSSB-mJ/view?usp=sharing

ANEXO E: VÍNCULOS CÓDIGO-CÓDIGO

https://drive.google.com/file/d/17FjBF8B_345araOaKMTxHUUmUxrjDpyf/view?usp=sharing

ANEXO F: LISTA DE CÓDIGOS Y DESCRIPCIONES

<https://drive.google.com/file/d/1DaDFszJ7VN0c7ChwclQdHxqKa666zd3o/view?usp=sharing>

ANEXO G: TRANSCRIPCIÓN DE ENTREVISTA DE LORENA

https://drive.google.com/file/d/15s9KyRd7SWy-wcXAU39_ZfbHwIsRMFJ0/view?usp=sharing

ANEXO H: TRANSCRIPCIÓN DE ENTREVISTA DE GENARO

https://drive.google.com/file/d/1444QJFNf_UVa9pw_x-9XHAL3_HEeozhr/view?usp=sharing

ANEXO I: TRANSCRIPCIÓN DE ENTREVISTA DE GRACIELA

<https://drive.google.com/file/d/1YmJNdnF3BJ4cVP15NjNvK3ftOOr50Uhi/view?usp=sharing>

ANEXO J: TRANSCRIPCIÓN DE ENTREVISTA DE VIRGINIA

<https://drive.google.com/file/d/1bxqLE4QkdX9wh6-bHoOVd70PnhQ1tUBq/view?usp=sharing>

ANEXO K: TRANSCRIPCIÓN DE ENTREVISTA DE CAMILA

https://drive.google.com/file/d/1y5IP7jFZE6IQkSL3e6ZJ3g5hbP_PVf_/view?usp=sharing

ANEXO L: TRANSCRIPCIÓN DE ENTREVISTA DE ALONDRA

<https://drive.google.com/file/d/1g6Zx-PICjuL-yiyimiTS7zoZGIRJMqylD/view?usp=sharing>

ANEXO M: TRANSCRIPCIÓN DE ENTREVISTA DE BELINDA

<https://drive.google.com/file/d/143anNPaZ031QfNdwYtpFC5xTMT7QLB-O/view?usp=sharing>

ANEXO N: TRANSCRIPCIÓN DE ENTREVISTA DE SIMÓN

<https://drive.google.com/file/d/1ISxurX8nAIaW1qq3I6utkcUAKbHefXX5/view?usp=sharing>

ANEXO O: TRANSCRIPCIÓN DE ENTREVISTA DE ERIKA

https://drive.google.com/file/d/1G2Gi2GRxQ1CfX5_GxAGOIy2gi7T4po0_/view?usp=sharing

ANEXO P: TRANSCRIPCIÓN DE ENTREVISTA DE NORA

<https://drive.google.com/file/d/1vep88ZI866Z0IVm0U9kkasdfxDaP0FDf/view?usp=sharing>

ANEXO Q: TRANSCRIPCIÓN DE ENTREVISTA DE GINA

<https://drive.google.com/file/d/1hGIRiwAsfYYCpSqCp9mR-9PGjwBq4Emy/view?usp=sharing>

ANEXO R: TRANSCRIPCIÓN DE ENTREVISTA DE NADIA

<https://drive.google.com/file/d/1zUn6TYmBbpFu48fkfLkx3lzKhnwAcD8o/view?usp=sharing>

ANEXO S: CONCENTRADO DE CITAS DE ENTREVISTAS

https://docs.google.com/spreadsheets/d/1QG3I_KR_0j2kiVv9uNpgbQxPp3CseZax/edit?usp=sharing&ouid=100406262656415828887&rtpof=true&sd=true

ANEXO T: MATRIZ DE FACTOR ROTADO

	FACTOR					
	1	2	3	4	5	6
ITEM057	,773					
ITEM058	,756					
ITEM056	,679					
ITEM055	,501	,411				
ITEM059	,473					
ITEM063		,623				
ITEM053		,612				
ITEM051		,561				
ITEM039		,560				
ITEM052		,544				
ITEM025		,419				
ITEM049		,407				
ITEM032		,406				
ITEM044		,369				
ITEM012			,672			
ITEM008			,561			
ITEM013		,369	,505			
ITEM022			,502			
ITEM048			-,447			
ITEM009			,438			

ITEM046	-,438		
ITEM015	,387		
ITEM041		,866	
ITEM040		,594	
ITEM061		,460	
ITEM060		,446	
ITEM001		,393	
ITEM054			,665
ITEM005			,624
ITEM027	,355		,502
ITEM021			,469
ITEM062			,454
ITEM026			,833
ITEM014			,568
ITEM030			,408

ANEXO U: CÓDIGOS DEL ANÁLISIS DE LAS ENTREVISTAS

(Elaboración propia, Atlas TI).

Totales		Totales		Totales	
● Confiabilidad	369	● Cambio	17	● Certeza	3
● Alteridad	300	● Certificación lengua	17	○ Complejidad	3
● Exclusión	227	● Metodología	17	● Cumplimiento	3
● Habilidades	216	● Ponderación	17	● Engaño	3
● Inclusión	208	● Posibilidad	16	● Esencia	3
● Cuidado alumno	201	● Variedad	16	● Especificaciones	3
● Validez	183	● Coherencia	15	● Incertidumbre	3
● Reflexión	175	● Esfuerzo alumno	15	● Inseguridad	3
● Validez Contenido	171	● Claridad	14	● Necesidad alumno	3
● Tareas	137	● Compensación	14	● Pensamiento	3
● Opinión	126	● Estructura	14	● Respeto	3
● Practicidad	120	● Evidencia	14	● Rigidez	3
● Gramática	102	● Material auténtico	14	● Selección texto	3
● Impacto	95	● Diseño original	13	● Sentimiento	3
● Dificultad	90	● Evaluación continua	13	● Trabajo	3
● Diálogo	86	● Necesidad	13	● Altruismo	2
● Vocabulario	83	● Planeación	13	● Atención	2
● Autoevaluación	82	● Conocimiento técnico	12	● Audacia	2
● Juicio alumno	82	● Interactividad	12	● Capacidad	2
● Adaptación	77	● Memoria	12	● Dosificación	2
● Colaboración	74	● Organización	12	● Estandarización	2
● Temática	73	● Referencia	12	● Formalidad examen	2
● Nivel de lengua	61	● Retroalimentación alumno	12	● Hoja de respuestas	2
● Preparación técnica	61	● Sensibilidad	12	● Influencia	2
● Intención	60	● Justificación	11	● Innovación	2
● Conocimiento	58	● Reto alumno	11	● Intuición	2
● Obligación	53	● Competencia comunicativa	10	● Marco de referencia	2
● Calificación	51	● Motivación	10	● Monetización	2
● Tecnicismo	51	● Empatía	9	● Oportunidad mejora	2
● Preferencia	50	● Use of English	9	● Sistematización	2
● Valoración negativa	49	● Comodidad	8	● Subjetividad	2
● Justicia	47	● Copiar pegar	8	● Suerte	2
● Experiencia	46	● Cultura	8	● Uniformidad	2
● Importancia	44	● Validez constructo	8	● Utilidad	2
● Práctica docente	44	● Calidad	7	● Validación	2
● Autenticidad	40	● Comparación	7	● Abuso de poder	1

● Estructura gramatical	39
● Valoración positiva	39
● Libro de texto	38
● Evaluación alternativa	37
● Desarrollo	35
● Urgencia	35
● Profesionalización	34
● Balance	30
● Práctica	30
● Uso de lengua	30
● Face validity	29
● Facilidad	29
● Formación alumno	29
● Mejora	29
● Política institucional	29
● Valoración	29
● Extensión examen	28
● Revisión	28
● Duda	27
● Autocrítica	26
● Impersonalidad	24
● Utopía	24
● Fairness	23
● Aprendizaje	21
● Contextualización	21
● Congruencia	19
● Objetividad	18

● Confiabilidad alumno	7
● Imposibilidad	7
● Incongruencia	7
● Limitación	7
● Obligación alumno	7
● Responsabilidad	7
● Seguimiento alumno	7
● Significatividad	7
● Banco de reactivos	6
● Competencia	6
● Concientización	6
● Inconformidad	6
● Desconfianza	5
● Interés alumno	5
● Reflexión alumno	5
● Aceptación	4
● Compromiso	4
● Constructo	4
● Crítica colegas	4
● Entendimiento	4
● Evaluación integrada	4
● Exigencia	4
● Inversión tiempo	4
● Pereza	4
● Problema	4
● Resignación	4
● Ayuda	3

● Aprendizaje autónomo	1
● Aprendizaje significativo	1
● Calibración	1
● Contacto hablantes	1
● Contradicción	1
● Credibilidad	1
● Delimitación	1
● Disculpa	1
● Dominio	1
● Estereotipos	1
● Ética profesional	1
● Evaluación formativa	1
● Expectativa	1
● Exposición L2	1
● Imagen	1
● Indecisión	1
● Interacción	1
● Interferencia L2	1
● Lingüística	1
● Originalidad	1
● Ortografía	1
● Precaución	1
● Prejuicio	1
● Profesionalismo	1
● Tecnología	1
● Teorización	1
● Tolerancia	1
● Washback	1