



3  
22in  
UNIVERSIDAD NACIONAL AUTONOMA  
DE MEXICO

FACULTAD DE FILOSOFIA Y LETRAS

" DISEÑO Y VALIDACION DE UN EXAMEN  
DE INGLES COMO LENGUA EXTRANJERA  
PARA ESTUDIANTES UNIVERSITARIOS "

T E S I S  
QUE PARA OBTENER EL TITULO DE  
LICENCIADO EN LENGUA Y LITERATURA  
MODERNAS INGLESAS  
P R E S E N T A:  
SARA EMILIA CHAVEZ VAZQUEZ

ASESORA DE TESIS:

MTRA. EMILIA REBORA TOGNO

TESIS CON  
FALLA DE ORIGEN

MEXICO, D.F.

1999

269704



Universidad Nacional  
Autónoma de México



**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.

*A mis padres, Sara y Carlos*

*Por su amor incondicional y su apoyo en todos los momentos de mi vida.*

*A mi esposo y compañero Marco Antonio*

*Por tu amor y el apoyo que siempre me has dado, no sólo como esposo, sino también como compañero de trabajo y colega con el que puedo compartir mis ideas y proyectos.*

*A mi hermano Carlos*

*Por tu cariño de hermano y tu ayuda para realizar las gráficas que parecían imposibles de hacer.*

*A mi hija Daphne*

*Por darme todos los días momentos inolvidables que guardo en mi corazón.*

## **Agradecimientos**

### **A la Mtra. Margarita Ordaz**

*Porque en estos años de intensa formación, su guía y consejo han dejado honda huella en mí, como profesora de inglés y como persona.*

*Por darme la oportunidad de compartir su experiencia y conocimiento trabajando a su lado en la Universidad Americana de Acapulco. Especialmente, gracias Margarita por el apoyo y cariño que siempre has tenido para mí.*

### **A la Mtra. Emilia Rébora Togo**

*Por su guía y apoyo en la realización de esta tesis. Especialmente agradezco la paciencia e interés que demostró en mi trabajo, a pesar del tiempo tan extenso que me tomó realizarlo.*

### **A las maestras integrantes del jurado**

**Lic. Lourdes Domínguez**

**Lic. Geraldine Gerling**

**Lic. Luz Aurora Félix**

*Por sus comentarios y sugerencias que ayudaron a aclarar mis ideas, y a mejorar este trabajo sustancialmente.*

### **Al Mtro. Joaquín Meza Coria**

*Por sus invaluable comentarios sobre este trabajo que reflejan su profundo conocimiento del ser humano y su amplia experiencia en el trabajo de investigación.*

*Gracias, Joaquín, porque el conversar e intercambiar ideas contigo siempre ha tenido la virtud de situarme y encontrar aspectos de mí misma que no conocía, y que han sido muy útiles en mi vida profesional y personal.*

### **A la Mtra. Alma Ortiz**

*Por su ayuda en el análisis de reactivos y su asesoría en la interpretación de los resultados obtenidos en las tres aplicaciones.*

### **A la Srita. Elba Nava Vázquez**

*Por su paciencia y dedicación en mi aprendizaje de los programas de cómputo utilizados en este trabajo.*

### **Al Profesor Salvador Becerra**

*Por su ayuda en la captura de los datos del análisis estadístico.*

**Diseño y validación de un examen de inglés  
como lengua extranjera para estudiantes universitarios.**

**Índice**

**I. La Universidad Americana de Acapulco y el Programa Lower**

A. Entorno institucional	p. 1
B. El Programa Lower	p. 3
C. El examen Lower como instrumento de evaluación :justificación teórica y práctica.	p. 10
D. Marco teórico del examen Lower	p. 12
1. Confiabilidad y validez	
2. Tendencias generales de diseño de exámenes de lengua extranjera según los diferentes enfoques	p. 24

**II. Proceso de diseño del examen**

A. Introducción	
1. Propósitos del examen Lower	p. 30
2. Proceso de diseño del examen Lower	p. 32
B. Descripción del examen	
1. Comprensión auditiva	p. 34
2. Gramática	p. 34
3. Vocabulario	p. 36
4. Comprensión de lectura	p. 37
5. Producción escrita	p. 37
6. Producción oral	p. 38
C. Cuadernillo de examen, instructivos y anexos	p. 39
D. Recursos humanos, materiales y financieros	p. 41

**III. Análisis de las tres primeras aplicaciones del examen Lower**

A. Introducción	p. 42
B. Descripción de las poblaciones	p. 46
1. 1a. aplicación	p. 46
2. 2a y 3a. aplicaciones	p. 46

### C. Reporte de resultados

- 1 Resultados *globales* p. 47
- 2 Resultados por sección p. 50

### IV. Conclusiones p. 68

- A Acerca de los resultados globales p. 68
- B Acerca de los resultados por sección p. 69
- C Consideraciones finales p. 82
  - 1. Respecto al examen Lower como instrumento de evaluación p. 83
  - 2 Respecto al Programa Lower p. 87

## Introducción.

El objetivo de este trabajo es presentar los resultados del análisis estadístico sobre el índice de facilidad y el análisis de reactivos de un examen de inglés como lengua extranjera que se diseñó en la Universidad Americana de Acapulco (UAA) durante el periodo comprendido de enero de 1993 a junio de 1994. El examen en cuestión recibió el nombre de **Lower**, término que se adoptó a partir del nivel terminal de la lengua extranjera propuesto en los programas de los cursos de la UAA, nivel definido en la *American Council for the Teaching of Foreign Languages Proficiency Guidelines* como *Intermediate Mid*.

Es conveniente aclarar que el término Lower no tiene relación alguna con el examen del mismo nombre que aplicó el Consejo Británico hace algunos años. En la opinión de autoridades y profesores que estuvimos involucrados en el proyecto de creación de un Centro de lenguas en la universidad mencionada, creemos que el término Lower define el nivel mínimo aceptable en los alumnos. futuros profesionistas. de la UAA.

El examen Lower es un examen de aprovechamiento que en un principio se aplicó a la Facultad de Contaduría, Administración e Informática y a la Facultad de Derecho con el fin de evaluar el desempeño de los alumnos en la lengua meta después de haber cursado un programa de inglés como lengua extranjera, programa que también se denominó Lower. Actualmente el Programa y el examen Lower únicamente se aplican en la Facultad de Contaduría, Administración e Informática debido a un cambio en el plan de estudios de la licenciatura en Derecho a partir de 1994, al que me referiré posteriormente.

El examen Lower como instrumento de evaluación ha sufrido modificaciones desde ese entonces a la fecha. Esos cambios han surgido de la necesidad de mejorar y en algunos casos, actualizar el instrumento de acuerdo a los resultados del análisis estadístico presentados en este trabajo y a mi experiencia y observación como profesora de inglés en la UAA.

El proceso de diseño del examen Lower se inició en enero de 1993 con la necesidad, ya contemplada desde la creación de la Dirección de Lenguas Extranjeras de la UAA, de contar con un examen que evaluara los conocimientos de los alumnos en la lengua extranjera al finalizar los cuatro cursos de inglés obligatorios. Por otra parte, este examen también cumpliría con la acreditación del requisito de titulación de la UAA mediante una calificación numérica.

Para tal fin, se instituyó un equipo de trabajo formado por la directora de la entonces Dirección de Lenguas Extranjeras (hoy Centro de Lenguas Extranjeras o CLE), la Mtra. Margarita Ordaz Mejía, y en el que yo participé como parte de mis funciones como secretaria académica.

La formulación de los objetivos del examen, el diseño de cada una de las secciones que lo componen y de sus anexos e instructivos quedó a cargo de la Mtra. Ordaz y la autora. Durante este proceso contamos con la ayuda de los profesores del 4º nivel de inglés, quienes nos apoyaron en la aplicación y calificación de las primeras pruebas de los reactivos de gramática y vocabulario, en el diseño de la sección de vocabulario especializado y en el primer piloteo del examen. Finalmente, el análisis estadístico, sus resultados y conclusiones, fueron responsabilidad de la autora del presente trabajo, cuyo propósito es presentar y analizar los resultados obtenidos durante este proceso, y de esta forma, establecer el grado de validez del examen Lower como instrumento de evaluación.

Sin embargo, al contemplar el examen Lower dentro del contexto para el que fue concebido, los resultados no sólo aportaron datos sobre el desempeño de los alumnos y su funcionamiento del instrumento de evaluación en sí, sino también plantearon interesantes cuestionamientos sobre el programa, los objetivos de enseñanza, los materiales didácticos y sobre la enseñanza misma. Por esta razón, también se incluyeron estos comentarios y reflexiones en las conclusiones de esta tesis.

En cuanto a la organización del presente trabajo, en el Capítulo I haré una descripción breve de la Universidad Americana de Acapulco, institución de educación superior y su Centro de Lenguas Extranjeras, organismo encargado de impartir las clases y evaluar el desempeño de sus alumnos. Comentaré además los aspectos teóricos más relevantes que fueron considerados por el equipo de trabajo en el diseño del examen y que a manera de marco teórico lo fundamentan y explican. Tales aspectos incluyen el tipo de examen diseñado de acuerdo a la función y a la metodología escogida, así como la importancia de los conceptos de confiabilidad y validez en un instrumento de evaluación.

El Capítulo II está orientado a la descripción del proceso de diseño de cada una de las secciones que conforman el examen, desde el análisis de los objetivos terminales de los cursos hasta la investigación sobre los formatos apropiados para evaluar tales objetivos y el diseño mismo del examen. De la misma manera incluiré una descripción detallada no sólo del examen, sino también de los anexos e instructivos con los que cuenta. Finalmente comentaré sobre los recursos materiales, financieros y humanos necesarios para la aplicación y calificación del examen.

En el Capítulo III haré una exposición comparativa de los resultados del análisis estadístico de las diferentes secciones del examen, incluyendo el índice de facilidad por reactivo y sección, y el índice de discriminación por reactivo en el caso de las secciones cuyo formato se preste idóneo para este tipo de análisis, como es el caso de las secciones de gramática, vocabulario y comprensión de lectura. En el caso de la sección de comprensión auditiva presentaré los resultados obtenidos sobre el índice de facilidad. Para la sección de producción escrita se incluye, además de los resultados del análisis de los trabajos de los alumnos, un breve comentario sobre la frecuencia de selección de los temas que eligieron los alumnos para trabajar. Finalmente, para la sección de producción oral se incluye tablas y gráficas que muestran comparativamente el desempeño de los alumnos en las tres diferentes aplicaciones.

Posterior al análisis estadístico, me permitiré hacer las recomendaciones pertinentes de acuerdo a los resultados obtenidos. Algunas estarán encaminadas a obtener resultados más

confiables en algunas secciones, mientras que otras recomendaciones estarán orientadas hacia otros aspectos .

En las conclusiones, se enfocará el examen desde otra perspectiva. A diferencia de las emendas en el capítulo anterior, mis recomendaciones no se centran en el análisis de reactivos, sino en el análisis de las secciones en cuanto al logro de sus objetivos y a la luz de algunas teorías sobre el diseño de exámenes. Mi objetivo en este punto es hacer una reflexión sobre la efectividad del examen Lower como instrumento de evaluación, y a manera de espejo establecer la relación de los resultados obtenidos en el análisis y los objetivos de enseñanza del Programa Lower.

Considero que si bien el trabajo que he realizado se ha circunscrito a un ámbito académico particular (los alumnos de ambas Facultades de la UAA), el marco teórico que lo fundamenta, las soluciones tomadas en cuanto a la elaboración de un examen, el análisis *porcionizado de los componentes* del instrumento de evaluación y la revisión crítica de la relación que guarda con los cursos que lo anteceden son, sin lugar a dudas, una fuente rica para todos los que estamos inmersos en la enseñanza de lenguas extranjeras en instituciones superiores.

Esta reflexión desde el punto de vista del profesor de inglés como lengua extranjera interesado en el proceso de evaluación y en el diseño de instrumentos válidos y a la vez confiables, ha sido muy importante para mi formación profesional en esta área. Especialmente me ha dado la oportunidad de poner en práctica los conocimientos adquiridos durante mis estudios de licenciatura en Lengua y Literatura Modernas Inglesas y para mi formación como profesora ha sido una experiencia enriquecedora que me ha abierto nuevas perspectivas para mi trabajo futuro.

A manera de ejemplo comentaré que en este trabajo únicamente se hizo un análisis estadístico de los resultados obtenidos en las tres primeras aplicaciones del examen. Sin embargo, en el curso de la investigación bibliográfica encontré algunos otros aspectos que

considero de suma importancia para el proceso de evaluación del instrumento mismo. Especialmente existen propuestas de investigación muy interesantes y útiles sobre la validez del instrumento. Estas propuestas son relativamente nuevas<sup>1</sup> y creo que serán de gran utilidad para todos aquéllos que nos interesamos en el campo de la evaluación y el diseño de exámenes, no solo desde el punto de vista del diseñador del instrumento, sino de desde el punto de vista de los seres humanos para los que un examen es diseñado y finalmente aplicado.

---

<sup>1</sup> James W. Popham, "The Validity of Score-based Inferences" en *Modern Educational Measurement: A Psychological Perspective*, p. 115-121.

## I. La Universidad Americana de Acapulco y el Programa Lower.

### A Entorno institucional.

Con el objeto de proporcionar un panorama del contexto para el cual el examen Lower fue elaborado, empezaré por explicar brevemente los orígenes del proyecto institucional de la Universidad Americana de Acapulco (UAA) y el lugar que ocupa la enseñanza del inglés como lengua extranjera dentro de la Universidad.

La UAA es una institución educativa fundada en agosto de 1992 que tiene como objetivo el impartir educación superior mediante la formación integral de sus estudiantes, procurando sensibilizarlos y capacitarlos para el ejercicio de su profesión.<sup>1</sup> Desde sus inicios, su objetivo fundamental ha sido la formación de profesionistas para los cuales se deberá propiciar el desarrollo integral en las áreas de conocimientos técnicos, culturales, sociales y humanos. Actualmente en la Universidad se imparten las licenciaturas de Contaduría, Administración, Informática, Derecho, Comunicación y Relaciones Públicas, Administración de Empresas Turísticas, Arquitectura, Psicología e Ingeniería en Computación.

Con este perfil en mente, la UAA ha diseñado sus programas y planes de estudios y ha instituido diferentes organismos encargados de promover y desarrollar sus objetivos. Uno de estos organismos fue la Dirección de Lenguas Extranjeras (DLE) que a partir de enero de 1994 dio origen al Centro de Lenguas Extranjeras o CLE, cuya finalidad es, entre otras cosas, el impartir clase de lengua extranjera a las diferentes facultades que integran la UAA. El CLE ha impartido desde sus inicios clases de inglés y francés como una herramienta necesaria que el alumno pueda utilizar para responder adecuadamente a las necesidades propias de su vida académica y posteriormente a las de su vida profesional.

Esto último necesariamente marcó la pauta para el diseño de los programas de cada licenciatura, ya que cada una de ellas necesita la lengua extranjera para propósitos diferentes. Por

---

<sup>1</sup> Universidad Americana de Acapulco. *Estatuto Orgánico*, p. 1.

esta razón llevamos a cabo reuniones de trabajo con los directores y secretarios académicos de las Facultades con el fin de hacer un análisis de las necesidades que tendrían los alumnos de las mismas en relación a la lengua extranjera. De la misma manera, revisamos los planes de estudio de algunas carreras, como Administración de Empresas Turísticas y Comunicación y Relaciones Públicas, carreras incorporadas a la Secretaría de Educación Pública, en las que la lengua extranjera es una asignatura curricular

Sin embargo, éste no fue el caso de las licenciaturas incorporadas a la UNAM. En estas licenciaturas la lengua extranjera no es una asignatura que forme parte de la currícula, sino que en algunas de ellas como en el caso de Contaduría, la comprensión de lectura de una lengua extranjera es solo un requisito para la titulación. En otras, como en el caso de Derecho, el requisito de titulación no existía al momento de diseño del examen. Sin embargo, por las razones anteriormente citadas, la UAA decidió incluir la lengua extranjera como asignatura obligatoria para todas las carreras. Como resultado de lo anterior, los planes de estudio de estas carreras, además de cumplir con los requerimientos de la Dirección de Incorporación y Revalidación de Estudios de la UNAM, incluyen el programa de lengua extranjera pertinente para cada una de las licenciaturas.

Como respuesta a este requerimiento curricular surgió el Programa Lower, que consta de cuatro cursos o niveles de inglés de 100 horas cada uno, es decir, que el curso tiene una duración total de 400 horas de clase repartidas en cuatro semestres. Este número de horas corresponden al nivel *Intermediate-Mid* de las *American Council for the Teaching of Foreign Languages (ACTFL) Proficiency Guidelines* de la Universidad de Michigan.<sup>2</sup>

En un principio el programa se impartió a la Facultad de Contaduría, Administración e Informática y a la Facultad de Derecho. Sin embargo, a partir de agosto de 1994, debido a un cambio de plan de estudios en la Facultad de Derecho, el objetivo de la asignatura de lengua extranjera se modificó a un curso de comprensión de lectura, por lo que en la actualidad el programa Lower sólo se aplica a los alumnos de la Facultad de Contaduría, Administración e Informática

---

Margarita Ordaz. *Proyecto de Reestructuración de la D.L.E.* p. p. 4-5

## B. El programa Lower.

Previo a la presentación de los objetivos de enseñanza del Programa Lower por nivel, es importante hacer una breve descripción del perfil de los alumnos que recibe este Centro y para los cuales fue diseñado este programa. Básicamente los alumnos se dividen en tres grupos:

- Los que provienen de las escuelas preparatorias incorporadas a la UNAM que han cursado 3 años mínimos de inglés de acuerdo a los programas vigentes.
- Los que provienen de las escuelas preparatorias incorporadas a la UAG que han cursado un año mínimo de inglés. Esta asignatura está orientada hacia la comprensión de lectura.
- Los que provienen de uno u otro sistema y que además han llevado cursos en forma particular o han estudiado en el extranjero.

Tenemos entonces una población heterogénea desde el punto de vista de sus conocimientos de inglés. Por tanto, la necesidad de un examen que clasificara a los estudiantes en niveles se hizo evidente. En atención al convenio de cooperación que el CELE de la UNAM ha mantenido con el Centro de Lenguas Extranjeras de la UAA desde su fundación, se solicitó su asesoría al respecto. Se sugirió la utilización de un examen de colocación diseñado en el propio CELE. Este examen consta de 120 reactivos de selección múltiple cuyo objetivo es determinar el nivel de conocimientos de la lengua extranjera con el fin de ubicar al estudiante en el nivel adecuado. De acuerdo a sus necesidades, el CLE dividió el examen en tres secciones que corresponden al 1º, 2º, y 3er. niveles respectivamente. Cada sección consta de 40 reactivos de los cuales el estudiante debe contestar 32 correctamente, es decir el 80%, para poder ingresar al siguiente nivel.

De esta forma aquel alumno cuyo examen de colocación indique que debe cursar los Niveles 3 y 4, tendrá automáticamente acreditados los Niveles 1 y 2 del programa Lower. Por otro lado, si los conocimientos del alumno no son suficientes y su examen de colocación indica que debe cursar desde el Nivel 1, dicho alumno tendrá que cursar los cuatro semestres completos de los que

---

consta el Programa. Al finalizar el curso, el alumno presentará el examen Lower, que junto con su promedio en el Nivel 4, decidirá sobre la acreditación del requisito establecido por la UAA.

El objetivo terminal que persigue el CLE queda expresado en el *Proyecto de creación de Lenguas Extranjeras* de la siguiente manera: “ La DLE impartirá clases de lenguas extranjeras a los estudiantes de la UAA con el objeto de proporcionarles habilidades de comunicación y de consulta bibliográfica que les permitirán desarrollarse profesionalmente.”<sup>3</sup>

Tomando en cuenta las ideas anteriormente expuestas, el CELE de la UNAM sugirió la utilización de sus propios programas de inglés para los primeros cuatro niveles. Tomando como base éstos últimos, el CLE hizo las modificaciones que consideró necesarias para adaptar los programas a la situación particular de Acapulco y de la UAA.

A partir de la idea general acerca de la metodología apropiada, se decidió aplicar los principios del enfoque comunicativo en la enseñanza de lenguas extranjeras ya que nos pareció idóneo para lograr la preparación requerida. Primero, porque dado que las necesidades primordiales de los estudiantes son la comunicación oral y escrita en su área de profesionalización en un puerto turístico como Acapulco, y que estas metas requieren de una competencia lingüística y comunicativa efectiva a nivel universitario, el enfoque comunicativo con sus aspectos funcionales, nocionales, gramaticales, socioculturales y pragmáticos, y sus técnicas de trabajo individuales y grupales, apoyaría metodológicamente el proceso adecuado de enseñanza-aprendizaje. Por otra parte, el principio organizador en funciones y nociones permitiría un sílabo en espiral, en el que una misma estructura puede ser utilizada en funciones lingüísticas diferentes y viceversa. Esto permite la revisión continua de funciones y estructuras proporcionando una práctica variada y rica donde el propósito es la comunicación significativa entre los interlocutores.

Para presentar los objetivos terminales de cada uno de los cursos en los que quedó integrado el Programa Lower, presento el siguiente cuadro *sinóptico* donde aparecen los niveles, el número de

---

<sup>3</sup> *Ibid.*, p. 14-15

horas/clase por semestre, las habilidades que cada curso privilegia y los objetivos terminales por nivel.

CURSO	HORAS/CLASE	OBJETIVO (S) TERMINAL (ES)
1er. Nivel, curso básico. Privilegia la comprensión de lectura sobre las otras habilidades: expresión oral, escrita y comprensión auditiva.	0 a 100 hrs.	<p>Sobre comprensión de lectura :</p> <p>Los alumnos desarrollarán estrategias de lectura que les permitirán comprender materiales auténticos como textos relacionados con el área de su profesión, artículos cortos sobre temas diversos, narraciones cortas y sencillas.</p> <p>Sobre las otras 3 habilidades :</p> <p>El estudiante alcanzará un uso limitado de la lengua. Será capaz de usar frases cortas y comprenderá estructuras básicas de la lengua extranjera, escribirá notas cortas y cartas informales breves.</p>
2o. Nivel, curso intermedio. Privilegia la comunicación oral sobre las otras habilidades.	100 a 200 hrs.	<p>Sobre las cuatro habilidades :</p> <p>El estudiante logrará emplear la lengua meta con efectividad en la comunicación básica oral y escrita en situaciones muy familiares. Podrá comprender y comunicar el sentido general en la mayoría de las situaciones con nativo hablantes acostumbrados al habla de no nativos.</p>
3er. Nivel, curso intermedio-avanzado. Privilegia la escritura sobre las otras habilidades.	200 a 300 hrs.	<p>El alumno alcanzará la preparación necesaria para llevar posteriormente cursos avanzados. Demostrará el sentido general en la mayoría de las situaciones con hablantes acostumbrados al habla de no nativos. En el campo de su profesión no encontrará dificultades con ningún tipo de textos escritos ni con estilos de habla de complejidad media.</p>
4o. Nivel, Curso intermedio-avanzado. Las 4 habilidades tienen la misma importancia.	300 a 400 hrs.	

Una vez definidos los objetivos terminales, describiré brevemente los contenidos de cada nivel.

Nivel 1. Como su objetivo lo indica, el Nivel 1 es, por un lado, un curso de tres habilidades, y por otro, incluye un curso de comprensión de lectura que por sí mismo podría ser un curso

independiente. El curso de comprensión de lectura responde a la necesidad de la consulta bibliográfica necesaria para la investigación documental que un estudiante de nivel superior tiene que realizar. De ahí que se haya incluido la comprensión de lectura desde el primer nivel de los cursos.

La hipótesis inicial de la que se partió proponía que dado que los alumnos eran falsos principiantes en inglés, la lectura les ayudaría a reconocer lo que ya sabían. En otras palabras, se propuso un proceso a la inversa de lo habitual, es decir, que en lugar de presentar el sistema gramatical de la lengua y posponer la lectura hasta que los estudiantes tuvieran mayores conocimientos de la lengua extranjera, se propuso iniciar con el discurso en la lengua meta (en este caso el discurso escrito) como medio para obtener información. En los cursos subsecuentes, estos alumnos poseerían un vocabulario más amplio en la lengua extranjera, reconocerían algunas estructuras ya vistas en la lectura y finalmente, dado que la lectura desarrolla otras habilidades y estrategias tanto de la lectura misma como de aprendizaje, éstas podrían ser aplicadas a otras habilidades, por ejemplo a la comprensión auditiva, habilidad privilegiada en el Nivel 2.

El programa de comprensión de lectura consta de ocho objetivos específicos que van desde la lectura global a la lectura específica y detallada, utilizando habilidades como la inferencia y la predicción, y estrategias como uso de apoyo visual, reconocimiento de términos transparentes en español y de las funciones retóricas más comunes.

Si tomamos en cuenta los cuatro niveles del programa, el Nivel 1 destaca porque es un programa completo, explícitamente desarrollado. Con esto último me refiero a que el programa de comprensión de lectura cuenta con los objetivos terminales y específicos, las estrategias de lectura, las funciones retóricas del discurso y los aspectos lingüísticos a considerar. Además incluye sugerencias de actividades y una propuesta clara de evaluación. En relación al programa sobre las otras tres habilidades (producción oral y escrita, y comprensión auditiva), éste incluye los objetivos terminales y específicos, los exponentes lingüísticos, el vocabulario, y las estructuras gramaticales pertinentes para cada una de las siguientes funciones:

- Expresar y responder a saludos (presente simple del verbo “to be”).
- Dar y pedir información personal: direcciones, preferencias y actividades diarias, sobre la familia, estudios (adjetivos posesivos).
- Expresar y responder a cortesías mínimas.
- Establecer metalenguaje en el salón de clases (auxiliares *do* y *does* como fórmula para preguntar o negar).
- Presentarse y presentar a otras personas.
- Invitar y responder a invitaciones.

*Nivel 2. Como los objetivos indican, el Nivel 2 es un curso global de cuatro habilidades en el que se privilegia la habilidad oral y la comprensión auditiva sobre las demás. Como en el nivel anterior, los objetivos específicos están ordenados en funciones y cada una de éstas viene acompañada de exponentes lingüísticos, vocabulario, estructuras y actividades sugeridas. Es importante mencionar que en este nivel se busca una expansión significativa del vocabulario. Sin embargo, como se mencionó anteriormente, no existe una propuesta de evaluación para este nivel. Las funciones incluidas son las siguientes.*

- Expresar y preguntar sobre el estado del tiempo.
- Hablar sobre actividades diarias. (presente simple del verbo “to be” y otros verbos, adverbios de frecuencia).
- Identificar a personas y objetos (presente simple y adjetivos demostrativos y posesivos).
- Describir objetos.
- Preguntar y dar información sobre lugares.
- Dar y pedir direcciones.
- Comprar y vender.
- Hablar acerca del presente (presente continuo).
- Hablar acerca de planes futuros, intenciones y promesas (*going to*).
- Hablar sobre experiencias en el pasado (pretérito simple).
- Hablar sobre el estado de salud.
- Hablar acerca de la temperatura.

- Expresar gustos y disgustos, preferencias.
- Hablar acerca de habilidades (modal *can* y adverbios de modo).
- Pedir un favor o permiso (modales *can* y *may*).
- Requerir algo.
- Hablar sobre obligaciones (modales *must* y *have to*).
- Decir un cumplido.
- Ordenar comida y bebidas.
- Preguntar sobre sentimientos, sensaciones.
- Emplear técnicas de conversación.

Nivel 3. Este nivel consta de 15 objetivos específicos y 3 objetivos de escritura para los cuales se creó un programa paralelo que se empezó a impartir de manera formal en agosto de 1994. Dicho programa está orientado a iniciar la redacción de composiciones y ensayos académicos cortos sobre temas de interés general o de su área de especialización. En cuanto a la lectura, se retoman en este nivel los dos últimos objetivos del programa del Nivel 1 (funciones retóricas y lectura crítica y detallada) y se inicia a los alumnos en la lectura de textos de su área de estudios. Los objetivos específicos son los siguientes:

- Presentarse y responder a presentaciones.
- Dar y pedir información
- Hablar acerca de experiencias pasadas (presente perfecto y adverbios de frecuencia).
- Dar y pedir opiniones.
- Narrar.
- Hablar acerca del futuro (*will* + infinitivo, presente continuo, *going to*).
- Expresar necesidades y deseos
- Comparar personas, lugares y cosas.
- Expresar un propósito (*going to*).
- Expresar posibilidad ( modales *may*, *might*.)
- Expresar obligación (modales *should*, *must* y *have to*.)
- Reportar órdenes y requerimientos.

- Pedir un favor, aceptar/ rechazar.
- Pedir a alguien que realice algo por nosotros ; aceptar/rechazar.
- Hablar acerca de posesiones.

Cabe hacer mención que el programa para este nivel no contempla explícitamente la evaluación. Con esto nos referimos a que el programa no cuenta con una propuesta para evaluar los contenidos del mismo.

Nivel 4. Este nivel es la última fase del Programa Lower, y por lo tanto, es también un curso de preparación para el examen del mismo nombre. En este nivel se incluyen funciones ya practicadas con anterioridad complementadas con nuevas y más complejas estructuras. Se privilegian la producción escrita y la lectura crítica y detallada relacionadas con el área de estudios, junto con vocabulario técnico propio de la carrera que el alumno haya elegido.

A partir de agosto 1994 se inició también la aplicación de un programa complementario de escritura diseñado por la autora. Este programa apoya esta habilidad con temas diversos, diferentes propósitos de escritura, funciones retóricas, marcadores del discurso y sugerencias de actividades.

Como he mencionado anteriormente, algunos aspectos del Programa Lower no están explícitamente desarrollados, otros se han ido desarrollando cada semestre de acuerdo a las necesidades que se han presentado. Este proceso es muy importante porque es parte de otro proceso continuo de actualización, valga la redundancia, y evaluación de los programas. A la fecha el programa Lower es una propuesta mínima, susceptible a modificaciones y adiciones según los requerimientos y las circunstancias cambiantes. Esta, que podríamos llamar "omisión", lejos de ser negativa, deja al profesor en libertad para organizar su curso y material de acuerdo a las necesidades de su grupo en particular y a su creatividad. Sin embargo, dentro de la perspectiva de la evaluación escolar es necesario contar con criterios específicos a los que el profesor y los alumnos puedan referirse, lo que representa un problema a resolver.

### C. El examen Lower como instrumento de evaluación : justificación teórica y práctica .

La evaluación como un análisis del proceso educativo plantea la necesidad de diferentes instrumentos que aporten información sobre los diversos aspectos que integran el fenómeno educativo. El programa que incluye los objetivos de enseñanza, es uno de estos aspectos junto con el logro de estos mismos objetivos por los estudiantes para los que fue concebido este programa. Aun cuando un examen no es el único medio por el que podemos obtener información sobre estos aspectos, sí es un instrumento por medio del cual tenemos un acercamiento a ambos, siempre y cuando éste cumpla con los requisitos para ser considerado como tal.

Este es el caso del examen Lower como instrumento de evaluación, ya que aporta información útil sobre el desempeño de los alumnos en la lengua extranjera tomando como criterios los objetivos terminales y específicos contenidos en el Programa Lower. Sin embargo, no sólo se obtiene información sobre el estudiante sino también sobre otros aspectos del proceso de enseñanza-aprendizaje, tales como el funcionamiento del programa en cuanto a sus objetivos tanto globales como específicos, la progresión de los mismos, las actividades de aprendizaje y los apoyos didácticos.

La evaluación comienza desde el inicio del proceso educativo con las metas y objetivos claramente definidos, seguido del diseño de instrumentos que permitan hacer evidente el logro o fracaso de estos objetivos. La información que se obtenga debe ser lo más confiable posible porque es a partir de ésta que se toman decisiones que afectarán de alguna manera al programa, a los cursos y al individuo. Por tanto, estas decisiones incidirán en todo el proceso de enseñanza aprendizaje.

La toma de decisiones es una parte importante del proceso de evaluación y está estrictamente relacionada con el propósito de la misma. La información obtenida por un proceso de evaluación es analizada e interpretada de acuerdo a los criterios establecidos, y a partir de estas conclusiones es que se toman decisiones sobre los diferentes aspectos evaluados. Para Hilda Taba, todo el fenómeno educativo puede ser evaluado "sus objetivos, alcances, calidad del personal docente, la preparación de los estudiantes, la importancia relativa de las diferentes materias, el grado en el que se cumplen

los objetivos, los medios de enseñanza, etc.”<sup>4</sup> Reiterando lo dicho, la evaluación sirve de base para tomar decisiones respecto a la pertinencia de los objetivos, a los criterios de contratación de los profesores y la formación que éstos requieran, el perfil de ingreso de los estudiantes, el lugar que ocupa la asignatura en relación a otras obligatorias u opcionales, la eficacia de los materiales y apoyos didácticos.

Si nos centramos en el alumno la evaluación influye sin lugar a dudas, en su futuro personal, por lo que es deseable que la información a partir de la cual se toman las decisiones sea lo más extensa y significativa posible. En este sentido, creemos que un examen único al final de cada ciclo escolar no es la forma más conveniente de evaluar el aprendizaje del alumno porque debido a las limitaciones del instrumento, tales como el tiempo del que se dispone y la extensión de la muestra que incluimos, probablemente no sea del todo representativa de las capacidades y conocimientos de los alumnos.

Sin embargo, las instancias educativas como la SEP y la UNAM exigen, al final de un año escolar una calificación numérica o alfabética que represente el grado de aprendizaje alcanzado por cada alumno. Esta representación numérica, la mayoría de las veces no aporta información relevante sobre los avances del alumno ni es válida en cuanto a que no incluye ni representa el dominio propuesto. Sin embargo, los alumnos y las instituciones educativas deben cumplir con el requisito de la calificación aprobatoria o acreditación.

Ante tal conflicto el CLE ha procurado establecer un balance entre ambos, donde la calificación obtenida por el alumno en el examen final es sólo uno de los variados instrumentos de los que dispone el profesor para otorgar la acreditación. Los instrumentos en cuestión incluyen el trabajo en clase, actividades especiales, pequeñas pruebas y trabajos que se promedian con el examen final. En este contexto, el examen Lower, evalúa el grado de dominio del alumno en las diferentes áreas de las que comprende el examen y certifica el logro de los objetivos del programa y es por lo tanto un examen de aprovechamiento que incluye las funciones y estrategias propuestas por el programa mismo.

---

<sup>4</sup> Hilda Taba en Morán Oviedo, *Op. Cit.*, p. 22

Para concluir, Morán Oviedo sostiene que :

La evaluación es un proceso complejo que empieza con la formulación de objetivos que involucra la evidencia de su logro, los procesos de interpretación para llegar al significado de esta evidencia y los juicios sobre las seguridades y las deficiencias de los estudiantes, y que finaliza con las decisiones acerca de los cambios y las mejoras que necesitan el plan de estudios y la enseñanza.<sup>5</sup>

#### D. Marco teórico del examen Lower.

De acuerdo al concepto de evaluación expuesto, el equipo de trabajo se dio a la tarea de diseñar un examen que se ajustara tanto a las necesidades de la institución, que serán presentadas en el siguiente capítulo, como a las características básicas de un buen instrumento de evaluación y que serán objeto de discusión en la presente sección . Entre estas consideraciones nos referiremos a las siguientes

- Consideraciones sobre el tipo de examen que se busca diseñar en cuanto a su uso y al público al que está dirigido.
- Consideraciones teóricas sobre el tipo de examen que se busca diseñar de acuerdo a la metodología de la enseñanza de lenguas extranjeras que el CLE ha adoptado.
- Consideraciones sobre la necesidad de que el instrumento sea válido y confiable.

De acuerdo a David Baker, todos los exámenes y sus resultados llevan como consecuencia la toma de decisiones sobre la persona que es examinada <sup>6</sup> y es deseable que así sea ; de otra manera no podría ser considerado un examen Por lo tanto, habría que decidir, en primer lugar, si es necesario aplicar un examen o no lo es. Si la decisión al respecto es afirmativa, entonces tendrá que especificarse cuál es la decisión que está en juego antes de iniciar el proceso de diseño del examen propiamente dicho

---

<sup>5</sup> Porfirio Moran Oviedo, "La evaluación de los aprendizajes", p 23

<sup>6</sup> David Baker, *Language Testing*, p 43.

En el caso del examen Lower, podría decirse que además de los propósitos que se discutirán mas adelante, existe otro mucho mas general que es el siguiente: un examen es siempre un medio mas económico (en términos del tiempo, recursos materiales y humanos) que sustituye a un proceso de toma de decisiones más complejo conocido como *criterion procedure*<sup>7</sup> o evaluación con respecto al criterio, y lo que intenta demostrar es el desempeño del estudiante en una situación hipotética pero que, de alguna manera, refleja la realidad. En este sentido el examen evalúa lo que realmente, valga la redundancia, busca evaluar y al mismo tiempo predice el desempeño del alumno en la situación real de comunicación. En otras palabras, es requisito del examen cumplir con dos cualidades vigentes para cualquier instrumento de evaluación : éste debe ser confiable y válido en cuanto a su contenido y a su capacidad, por así llamarla, de predicción .

### 1. Confiabilidad y validez.

La preocupación por diseñar un examen que sea a la vez confiable y válido es evidente en los autores que han escrito sobre la elaboración y análisis de pruebas y exámenes. Además de David Baker citado anteriormente, se encuentran Alan Davies, John Oller y John B. Heaton, de gran renombre en el campo de la evaluación.

Todos ellos coinciden con Baker en que el examen es un sustituto de otro procedimiento más complejo por medio del cual se obtiene evidencias sobre los conocimientos y habilidades de los estudiantes en situaciones determinadas por los objetivos específicos de un curso. Reiterando lo dicho anteriormente, la evaluación es un proceso de dimensiones no sólo personales, sino sociales en un sentido mucho más amplio y trascendente.

Dada la importancia de las decisiones involucradas, es preocupación del diseñador que su instrumento sea válido y confiable asegurando de esta manera dos aspectos muy importantes para la toma de decisiones :

- Que el examen obtenga efectivamente la evidencia de las habilidades y/o conocimientos adquiridos que se propone evaluar, es decir que sea válido.

---

<sup>7</sup> *Ibid* , p. p 44-45.

- Que los resultados obtenidos sean producto de lo que queremos evaluar y que sean consistentes en términos de objetividad, es decir que sea confiable.

A través de un proceso de análisis, el diseñador examina, valga la redundancia, su instrumento de evaluación. Este proceso de análisis comienza en el diseño y continúa una vez que el examen ha sido aplicado a la población para la cual fue diseñado y sus resultados han sido cuantificados e interpretados. Como se puede observar, este es un proceso continuo, como la evaluación misma, por medio del cual se obtendrá información útil para el diseñador en relación a su instrumento.

Con el fin de explicar ambos, confiabilidad y validez, y mostrar de qué manera influyen en la situación y resultado de un examen, se hará una presentación de ambos conceptos, se definirán los factores que pueden influir negativamente en ellos y, de manera general, se propondrán alternativas para reducir al mínimo estos factores.

Tradicionalmente se ha enfatizado las diferencias entre ambos conceptos, confiabilidad y validez, considerándolos como aspectos distintos, incluso opuestos; y donde la permanencia de uno de ellos necesariamente conlleva la exclusión del otro. Esta que podríamos llamar *tensión* entre uno y otro queda claramente expresada en el trabajo de Alan Davies<sup>8</sup> refiriéndose a los nuevos tipos de exámenes (en oposición a los exámenes tradicionales), en donde declara que *“the only criticism we would make on the theoretical side is that of doubtful reliability: this (...) may be unavoidable in order to retain validity”*.

De acuerdo a la cita anterior, vemos claramente el conflicto o tensión entre ambos conceptos que se podría explicar de la siguiente manera: si por cuestiones de validez se procurara incluir en el examen tareas lo más parecidas posibles a una situación real, mismas que están incluidas dentro de los objetivos de enseñanza, estas tareas no siempre serían cuantificables desde el punto de vista psicométrico y objetivo en donde sólo es aceptable una respuesta correcta, sino que se necesitaría otro tipo de criterio para evaluar la tarea realizada. Si tomamos en cuenta que esta es

---

<sup>8</sup> Alan Davies. “The Construction of Language Tests”, p.54

una preocupación no sólo del enfoque psicométrico sino de la mayoría de los enfoques, estamos entonces ante un dilema que Charles Alderson define de la siguiente manera: *"The problem for most language testers is that in order to maximize reliability it is often necessary to reduce validity. The format might be not so valid to measure of one's ability to use the language"*<sup>9</sup> Esta preocupación es evidente al iniciar una investigación al respecto en donde autores importantes como los que citaré a continuación exponen su punto de vista sobre el particular. Según Davies,

*A reliable test possesses consistency of results. An inconsistent test would give meaningless, random results. Before looking at the meaning of results it is important to ensure that they are reliable. Unreliable results can have no meaning apart from their own randomness. The implication here is that although validity is far more important than reliability, [...], it is essential to establish reliability first. Otherwise there is no point in considering validity.*<sup>10</sup>

Si se toma en cuenta lo que dice Davies, veremos que ambas cualidades son necesarias en un examen, y que la validez depende en gran parte, de establecer un grado de confiabilidad al mismo

Heaton comparte la posición anterior al declarar lo siguiente: *"Reliability is a necessary characteristic of any good test: for it to be valid at all, a test must first be reliable as a measuring instrument. [...] In order to be reliable, a test, must be consistent in its measurements."*<sup>11</sup> Al respecto Bachmann opina que: *When we increase the reliability of our measures, we are also satisfying a necessary condition for validity. In order for a test score to be valid, it must be reliable.*<sup>12</sup>

En las citas anteriores destaca el hecho que, al tratar de definir el concepto de confiabilidad, los autores citados mencionan sin excepción la validez. Aunque pareciera que estos conceptos son mutuamente excluyentes, el uno no existe sin el otro y al establecer cualquiera de ellos

<sup>9</sup> Charles Alderson *et al.*, *Language Test Construction and Evaluation*, p.187.

<sup>10</sup> Davies. *Op Cit*, p 57.

<sup>11</sup> John B Heaton. "Criteria and Types of Tests", p. 155.

<sup>12</sup> Lyle F Bachmann, *Fundamental Considerations in Language Testing*, p. 161.

necesariamente tendremos que referirnos al otro, como veremos en las siguientes afirmaciones en donde existe un consenso de los autores citados :

- La confiabilidad se puede traducir en consistencia de resultados.
- La confiabilidad es un punto de partida para establecer la validez de un examen.
- La validez y la confiabilidad se interrelacionan.

Tomando como puntos de partida las afirmaciones anteriores, procederemos a analizar cada una de ellas. Primeramente vemos que a través de los años la confiabilidad se ha utilizado como consistencia en los resultados, consistencia que se manifiesta a través del tiempo. De acuerdo a esto, los resultados de un examen confiable no variarán si éste es aplicado a los mismos alumnos en diferentes ocasiones. W. James Popham define esta consistencia como *estabilidad*.<sup>13</sup> Sin embargo la confiabilidad no sólo puede ser considerada como estabilidad en términos de consistencia de resultados. Podría decirse que esta consistencia es producto de varios aspectos que trabajan de manera conjunta para establecer cierto grado de confiabilidad del examen. Según el mismo autor, la consistencia puede ser interna o externa, entendiéndose la primera como la consistencia entre los diferentes componentes del examen (determinada por las fórmulas de Spearman-Brown y el método de Kuder-Richardson,<sup>14</sup>) y la segunda como la comparación entre los resultados del examen aplicado en dos diferentes ocasiones (pudiendo utilizarse dos versiones diferentes del mismo examen) o la comparación de los resultados del examen con otros resultados de evaluaciones distintas al mismo, como pudieran ser los resultados que cada profesor tiene de sus alumnos.<sup>15</sup>

En cualesquiera de los métodos de enseñanza de lenguas, la preocupación sigue siendo la misma. asegurar que el instrumento esté midiendo de manera efectiva lo que se desea medir, es decir que si queremos evaluar la capacidad del alumno para obtener la idea general de un texto auditivo sea exactamente eso lo que se evalúa y que el resultado no dependa de otros factores como pudieran ser la ambigüedad en las instrucciones del examen o la subjetividad de la persona que esta emitiendo un juicio de valor sobre el desempeño del alumno.

---

<sup>13</sup> W. James Popham. *Modern Educational Measurement*, p. 122.

<sup>14</sup> *Ibid.*, p. 132-139.

<sup>15</sup> *Ibid.*

Es aquí precisamente en donde se encuentra el conflicto enunciado por Davies y Alderson: si se quiere establecer cierto grado de objetividad, se dará preferencia a formatos que nos aseguren que, a pesar de que haya diferentes examinadores, el resultado del candidato no se vea afectado por esta circunstancia. En una primera instancia, se preferirá un examen de gramática (posiblemente de selección múltiple) sobre un examen de expresión más libre como pudiera ser un examen de producción escrita por el simple hecho de que presenta mayores dificultades para asegurar el mismo resultado. El problema entonces es el siguiente: “*Objective tests and formats do not [always] reflect classwork.*”<sup>16</sup>

Al referirnos al trabajo de clase y por ende a los objetivos de enseñanza, estamos hablando de la validez, entendiéndose ésta como la correspondencia entre estos objetivos y la valoración que hace de ellos el instrumento.<sup>17</sup> Si la confiabilidad asegura la consistencia de resultados, la validez asegura la significación de los resultados,<sup>18</sup> entendiéndose ésta como la congruencia entre los objetivos del curso y el trabajo de clase. La pertinencia y significación de los resultados será sumamente importante porque es a partir de la validez que el examinador puede emitir un juicio en términos de lo que el candidato puede o no puede hacer con la lengua en relación a un criterio establecido. Sin embargo, a pesar de la importancia de este concepto y en contraste con la confiabilidad, existen pocos estudios al respecto de la validez, ya que se ha considerado suficiente el incluir los objetivos de enseñanza en el examen y describir el proceso por el medio del cual se llevó a cabo la selección, y de este modo, resolver el problema de la validez de contenido.<sup>19</sup> Al referirnos a este tipo de validez, consideramos importante mencionar los diferentes aspectos que toma en cuenta la validez, y sobre todo en el tipo de evidencia que cada uno aporta.<sup>20</sup>

Bachmann y Samuel Messick opinan en el sentido de que no podemos hablar de la validez de un examen. De acuerdo a Messick<sup>21</sup>, “*one validates, not a test, but an interpretation of data arising from a specific procedure*”. Por tanto para este autor lo importante son las interpretaciones

<sup>16</sup> Davies. *Op Cit*, p 51

<sup>17</sup> Robert Woods. *Assessment and Testing*, p. 147

<sup>18</sup> Davies. *Op Cit*, p 58

<sup>19</sup> Popham *Op cit.*, p. 99

<sup>20</sup> Bachmann. *Op Cit.*, p. 237

<sup>21</sup> Messick en Popham, *Op Cit*, p 94.

que se hagan a partir de la evidencia obtenida. Por su parte Bachmann agrega que la validez *"is not simply a function of the content and procedures of the test itself(...). Validation must therefore consider, in addition to the test's content and method, how test takers perform"*<sup>22</sup> Estas dos citas orientan entonces la investigación sobre la validez hacia la interpretación de datos o evidencia obtenidos por métodos y para propósitos diferentes, incluyendo la investigación sobre el proceso del candidato al momento del examen. Finalmente Bachmann afirma que *"it is only through the collection and interpretation of all relevant types of information that validity can be demonstrated"*<sup>23</sup> Surge entonces la pregunta sobre cuáles son los diferentes tipos de información que puede obtenerse en una investigación sobre la validez. Popham propone tres tipos de evidencia de acuerdo a la información que aporta cada una de ellas<sup>24</sup> :

■ **Evidencia de validez con respecto al contenido.** Como indica la definición de los *Standards for Educational and Psychological Testing* citada por Popham,<sup>25</sup> esta evidencia demuestra el grado en que los reactivos, tareas o preguntas del examen son representativas del universo definido o dominio. A diferencia de la validez aparente o *face validity* en donde la evidencia es aportada por personas que no necesariamente son expertos en el área, la evidencia de validez con respecto al contenido es obtenida mediante la opinión de expertos en la materia e inicia durante el proceso de diseño del examen y continúa después de que el examen ya ha sido aplicado. Para reunir la evidencia mencionada, se requiere del diseño de cuestionarios que, como ya se mencionó anteriormente, se aplicarán a expertos del área que se desea evaluar.

Es importante mencionar que aunque Popham califica la validez aparente como *"superficial"*<sup>26</sup>, algunos otros autores como Charles Alderson, Carolina Clapham y Dianne Wall opinan que este tipo de validez incluye un juicio intuitivo y global que se refiere al examen en su totalidad, aunque a veces también incluye comentarios sobre la ambigüedad de las instrucciones o sobre los tiempos dedicados a cada sección o tarea.<sup>27</sup> En relación a este tipo de validez, los autores

---

<sup>22</sup> Bachmann. *Op Cit*, p. 237.

<sup>23</sup> *Id*

<sup>24</sup> Popham, *Op Cit*, p p. 93-110

<sup>25</sup> *Ibid*, p 97

<sup>26</sup> *Id*

<sup>27</sup> Alderson *et al.* *Op Cit*. p p 172-173

mencionados afirman que es importante que el examen parezca aceptable a los usuarios, porque de esta impresión se derivará la actitud que los candidatos tomen con respecto al examen *“tests that do not appear to be valid to users may not be taken seriously for their given purpose”*<sup>28</sup>

- **Evidencia de validez con respecto al criterio.** Como su nombre lo indica., esta evidencia demuestra que los resultados del examen se relacionan sistemáticamente con el desempeño del candidato en relación a un criterio establecido. En este tipo de evidencia, se incluye la evidencia sobre la validez de predicción y sobre la validez concurrente, cuya diferencia es, principalmente, el intervalo de tiempo entre la aplicación del examen y la evaluación de alguna otra actividad que pueda proveer de la evidencia necesaria para comparar los resultados.
- **Evidencia de validez con respecto a la construcción,** cuya evidencia se enfoca principalmente a la puntuación obtenida en el examen como una medida de una característica psicológica de interés. En el caso de un examen de lengua, un ejemplo de esta evidencia son las diferentes funciones como objetivos incluidos en los programas de los cursos. Dada la diversidad de evidencias dependiendo del criterio que se desea evaluar, no existe un solo método para la obtención de estos datos, aunque pueden clasificarse en alguno de estos tipos de estudios: estudios de **intervención**, en donde la preocupación principal es demostrar que los candidatos responden diferentemente al recibir la instrucción adecuada; estudios de **diferenciación de la población**, en donde el énfasis radica en la diferencia de resultados obtenidos por individuos representativos de diferentes poblaciones; y finalmente, estudios sobre **medidas** proporcionadas por otros instrumentos de evaluación.

Como es evidente, en las propuestas de Popham, se insiste en la necesidad de que se hagan estudios sobre la validez de un instrumento de evaluación tal como se hace con la confiabilidad.

A pesar de que al concepto de validez no se le ha sistematizado como a su contraparte, algunos autores opinan que la validez es más importante que la confiabilidad. En este sentido Woods opina que *“Validity is ultimately more important than reliability but any attempt at*

---

<sup>28</sup> *Ibid.*, p. 173

*validation depends crucially on the reliability of observations. If these cannot be trusted, a misleading judgement concerning validity is likely to be reached*”<sup>29</sup> La importancia de la cita anterior, es que coloca a cada concepto en su justa dimensión y lugar. Es a partir de los resultados del examen, que el examinador emite un juicio estimativo. Por lo tanto, habrá que asegurar que los resultados a partir de los cuales se harán inferencias, sean confiables: *“A test is merely a tool we use in order to make inferences”*”<sup>30</sup>

Recordemos que Bachmann comparte la posición de Popham al declarar que *“In order for a test score to be valid, it must be reliable”*.<sup>31</sup> Bachmann propone entonces que la confiabilidad y la validez no son dos conceptos que se oponen el uno al otro; por el contrario, se complementan. Ambos intentan aclarar una preocupación en la medición: identificar, prever y controlar los factores que afectan las puntuaciones en los exámenes y los efectos que pudiesen tener sobre ellas. El autor sugiere que ambos conceptos deben cumplir con dos requisitos básicos en cualquier instrumento de evaluación:

- Minimizar los efectos del error en la medición.
- Maximizar los efectos de las habilidades lingüísticas que queremos evaluar.

Para concluir, el trabajo de investigación sobre la confiabilidad y la validez debe orientarse hacia el establecimiento de ambas, que, como ya hemos explicado, no son mutuamente excluyentes, por el contrario se complementan. Es importante identificar las fuentes de error en los diferentes tipos de evidencia que se obtienen a partir de estudios diferentes y estimar la magnitud de sus efectos sobre los resultados del examen.

En cuanto a la confiabilidad se refiere, Bachmann, Davies y Hughes coinciden en que los factores que se mencionarán a continuación afectan de alguna manera el grado de confiabilidad del examen. Algunos de ellos como la motivación, la fatiga, el estado de salud y la habilidad del candidato para contestar ciertos formatos del examen no pueden ser controlados por el diseñador. Sin embargo, existen otros factores que sí pueden verse. La confiabilidad establece la

---

<sup>29</sup> Woods. *Op Cit*, p. 147

<sup>30</sup> Popham. *Op Cit*, p. 94.

<sup>31</sup> Bachmann. *Op Cit*, p. 160.

independencia del desempeño del candidato de estos factores no relacionados con su conocimiento y uso de la lengua en este caso. Por lo tanto, la tarea a realizar será minimizar el efecto de éstos sobre los resultados del examen

**En cuanto al diseño del examen :**

- Definir claramente los objetivos del examen y los alcances de las tareas y reducir de esta manera la posibilidad de que el alumno interprete erróneamente o se confunda y no sepa qué se espera exactamente de él.
- En consecuencia, diseñar reactivos y tareas que no se presten a ambigüedades y de los cuales se obtenga suficientes muestras de lo que queremos evaluar. Es importante, sin embargo, que la muestra no pretenda ser tan extensa que cause a los estudiantes o éstos pierdan el interés dándose cuenta de que lo que están contestando es poco representativo de sus habilidades.
- Las instrucciones deben ser claras y explícitas.
- Incluir reactivos que permitan ser evaluados lo más objetivamente posible por varios examinadores.
- Asegurarse que la impresión del examen escrito sea de buena calidad, legible y ordenada.

**En cuanto a la aplicación y administración del examen :**

- Proporcionar las condiciones físicas y ambientales necesarias para la aplicación uniforme del examen, es decir, que se aplique de manera similar en circunstancias similares.

**En cuanto a la calificación de los exámenes :**

- Identificar a los candidatos por un número, y no por sus nombres
- Proporcionar a los examinadores una clave de respuestas
- Entrenar a los examinadores.
- Pedir que dos o más examinadores revisen cada examen.

**En cuanto a los candidatos :**

- Los candidatos deben conocer la información necesaria sobre el examen : objetivos, secciones y/o habilidades que se van a evaluar, el mínimo necesario para acreditar.

- Los candidatos deben estar familiarizados con los posibles formatos del examen.

Desde el punto de vista de la validez, ésta intenta especificar y diseñar los reactivos de tal modo que se asegure que el candidato utilice las habilidades sobre las que el diseñador se ha propuesto obtener evidencia. Esto nos remite necesariamente al programa y a los objetivos de enseñanza, siendo éstos últimos los que definirán el contenido tratándose de un examen de aprovechamiento o avance, como en el caso del Lower. Por lo tanto, siguiendo el procedimiento sugerido por Popham, la investigación sobre la validez debe aplicar dos estrategias distintas :

- Primero, los intentos de definir e incluir lo que será el contenido del examen debe hacerse durante el proceso de diseño del examen. Para este propósito, se requerirá la ayuda de colegas expertos con el fin de asegurar que el contenido del examen represente satisfactoriamente el contenido del curso <sup>32</sup>
- Segundo, llevar a cabo procedimientos "*post facto*" una vez que el examen ha sido aplicado para obtener evidencia de la validez del contenido en dos planos, tanto si cada reactivo representa el contenido de interés, es decir del que queremos obtener la evidencia, como en qué medida estos contenidos han sido medidos por los componentes del examen.<sup>33</sup> Reuniendo la evidencia anterior, se intenta sistematizar y cuantificar datos que usualmente no era preocupación de los diseñadores describir, como los datos aportados por los estudios sobre la confiabilidad.

Habiendo discutido ambos conceptos, comentaremos que en relación al examen Lower se tomaron en cuenta ambas cualidades y sugerencias de la manera que especificamos a continuación. En primer lugar, desde sus inicios fue preocupación de los diseñadores establecer un examen de aprovechamiento confiable y que nos permitiera obtener evidencias válidas sobre el desempeño de los candidatos. Esta preocupación determinó el tipo de examen que se buscaba diseñar. Existió un consenso en relación a que el objetivo central del examen era evaluar aquellos objetivos incluidos en los cursos, y establecer así un grado de dominio al que los candidatos deberían aproximarse.

---

<sup>32</sup> Popham. *Op Cit* , p. 98.

<sup>33</sup> *Ibid* . p. 101.

Como ya se ha mencionado en otras secciones del presente trabajo, iniciamos el proceso de diseño con un análisis tanto de los objetivos como de las actividades y evaluaciones periódicas de los profesores de todos los niveles del programa Lower. De esta manera, pretendimos establecer cierto grado de validez, desde el punto de vista de los contenidos y que al mismo tiempo, fuera congruente con la metodología de enseñanza del CLE, es decir con el enfoque comunicativo.

Sin embargo, también discutimos el problema de la confiabilidad. Nos propusimos que, aunque no era nuestro objetivo diseñar un examen del tipo psicométrico y objetivo, podríamos utilizar algunas ventajas de tales formatos, como en el caso de la sección de *gramática* y *vocabulario*. Este formato nos ofreció la ventaja de la economía de medios, al permitir que se incluyera mayor número de funciones comunicativas y además establecía cierto grado de *objetividad al examen*. A pesar de lo anterior, también buscamos incluir tareas que, aunque no podrían ser calificadas de objetivas en una primera instancia, si eran parte de lo que los estudiantes habían trabajado en clase y que estaban contempladas en los programas. Tal es el caso del ensayo académico y de la sección de *producción oral*, en donde existe una libertad mayor para utilizar la lengua con un propósito específico, pero de manera tal que el estudiante pudiera hacer uso de su conocimiento de la lengua y de sus estrategias de comunicación para transmitir su mensaje.

De la misma manera en cuanto a confiabilidad se refiere cuidamos factores tales como la suficiencia en el número de reactivos para proporcionar una muestra fehaciente; especificamos las operaciones e instrucciones eliminando imprecisiones y ambigüedades. Finalmente la aplicación del examen se programó de manera cuidadosa, *entrenando a los examinadores* en juntas programadas con este propósito. Cabe hacer mención que el análisis estadístico de la confiabilidad quedó pendiente en el momento del diseño; después de la primera aplicación, comencé con este análisis cuyos resultados y conclusiones presentaré en el siguiente capítulo.

## 2. Tendencias generales de diseño de exámenes de lengua extranjera según los diferentes enfoques.

En relación al tipo de examen según la metodología adoptada, se mencionarán cuales han sido las tendencias generales de diseño de exámenes en el campo de la enseñanza de lenguas extranjeras para después referirnos al diseño específico del examen Lower.

**Enfoque tradicional.** Este enfoque se basa en el concepto de la lengua como un fin en sí misma. Existe principalmente en forma de texto escrito, por lo que fue muy utilizado en la enseñanza de lenguas clásicas. No cuenta con una teoría psicológica de base, pero sí se apoya en la teoría tradicionalista de la enseñanza de lenguas<sup>34</sup> El método de enseñanza es la traducción, de manera que los estudiantes debían traducir los documentos escritos, de manera literal. El tipo de examen aplicado incluía traducción, listas de vocabulario, inflexiones de sustantivos y conjugaciones de verbos. Las necesidades o intereses de los alumnos no eran el punto focal de la enseñanza.

**Enfoque psicométrico/ estructuralista.** Este enfoque estuvo en boga en los años sesenta, y tiene sus raíces en dos concepciones filosóficas distintas. La primera, tiene que ver con el concepto de aprendizaje desde el punto de vista de la psicología conductista. La teoría conductista o *behaviorist* establece que el aprendizaje es formación de hábitos y respuestas condicionadas al medio en el que nos desenvolvemos. Incluye además la modificación de conductas consideradas erróneas o no deseables por otras aceptables. La lengua como medio de expresión trabaja mediante un mecanismo de estímulo respuesta, en donde el contexto no es necesario para producir respuestas gramaticalmente correctas. El análisis de la lengua se hace dividiendo sus componentes desde los sonidos hasta la formación de oraciones sintácticamente correctas. Por lo que, para enseñar una lengua es necesario dividirla en partes. El profesor ofrece al alumno el estímulo que demande una respuesta correcta por parte del alumno. Esta respuesta (que debe ser observable), demostrará si se crean los hábitos correctos o no.

---

<sup>34</sup> Davies, *Op. Cit.*, p. 40

En cuanto al tratamiento de errores, las respuestas correctas son reforzadas, mientras que las incorrectas tratan de ser erradicadas totalmente.<sup>35</sup> Por lo tanto, un buen examen de esta naturaleza examina cada uno de los componentes de la lengua por separado (las más de las veces gramaticales) mostrando con esto la habilidad del candidato para manejar este aspecto de la lengua. Un típico examen psicométrico incluye ejercicios de selección múltiple, llenar espacios y tal vez completar diálogos.

La ventaja de estos exámenes es que son fácilmente cuantificables, característica que en términos de confiabilidad eran, y siguen siendo, muy deseables. Sin embargo, algunos autores como Baker han mencionado algunas desventajas. En primer lugar, aunque no cuestiona la pertinencia del análisis de la lengua en partes y segmentos, sí cuestiona el problema de dividir la lengua para medir cada uno de sus componentes por separado en una situación de examen. Según sus propias palabras, la lengua es presentada "*unnatural[ly] and unlanguage-like*"<sup>36</sup>, es decir que no presentan la lengua tal y como ocurre en la vida real, o al menos lo más parecida a una situación real, con un contexto, una función y un propósito definido. De acuerdo a Morrow es evidente que el principal problema de este análisis atomístico es la habilidad del que aprende la lengua para sintetizar<sup>37</sup>, capacidad no siempre presente, y que los teóricos de este enfoque daban por hecho.

**Enfoque transformacional.** Este enfoque se basa en la teoría cognoscitiva del aprendizaje de la lengua, que se apoya en el reconocimiento de los componentes y el significado de las palabras. El aprendizaje formal y la lengua son independientes uno del otro porque nunca dejamos de aprender la lengua materna. La lengua es un medio creativo de expresión y está basada en reglas muy específicas compartidas por los nativos hablantes. La teoría Generativa Transformacional de Chomsky es la columna vertebral de este enfoque desde el punto de vista lingüístico. Las especificaciones gramaticales son necesarias e involucran la presentación de varios ejemplos con el objeto de hacer explícitas las estructuras profundas que subyacen a las estructuras superficiales. La lengua como un sistema fijo es el objeto de estudio de este enfoque, y deja de lado "*such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and*

<sup>35</sup> Cfr. Keith Morrow, "Communicative Language Testing", *passim*

<sup>36</sup> Baker, *Op Cit*, p. 65

<sup>37</sup> Morrow, *Op Cit*, p. 145

*interest, and errors (random or characteristic)* <sup>38</sup> Este concepto ha servido de base para una gran cantidad de exámenes de lengua. Según Morrow las condiciones mencionadas anteriormente pueden ser gramaticalmente correctas o no, pero no podemos dudar que existen en cualquier situación de comunicación.

**Enfoque pragmático/integrativo.** Como resultado de los estudios e investigaciones sobre la naturaleza de la lengua y los mecanismos de su adquisición y aprendizaje en la década de los años sesenta, se promovió un cambio necesario en el concepto de lengua, de enseñanza de lenguas extranjeras y segunda lengua, de programas, materiales y evaluación. Aparecieron los sílabos funcionales y nocionales que cuestionan el orden y la selección de los componentes de la lengua en base únicamente de las estructuras sintácticas. Se propuso un sílabo ordenado en nociones y funciones donde la lengua siempre tiene una intención<sup>39</sup> como instrumento de comunicación.

Como es de suponer el interés en las funciones y nociones de la lengua trajo como resultado un cambio: de ser éste considerado un sistema hermético y de reglas inalterables, la lengua pasó a tener dimensiones de fenómeno psicológico y social. Al tomar en cuenta las necesidades e intereses del alumno de una lengua extranjera, aparecieron los cursos para propósitos específicos, o ESP, en los cuales los objetivos están orientados a que el alumno desarrolle ciertas habilidades que le permitan resolver situaciones propias de su área de estudios, trabajo o interés. Necesariamente, todos estos cambios en la metodología incidieron en un nuevo planteamiento de la evaluación y el diseño de exámenes

Carroll (1961) es el primero en hacer la distinción entre los términos "*discrete-point test*" e "*integrative test*" lo que permite distinguir entre ambas posturas. La primera "*is one that attempts to focus attention on one point of grammar at a time and only one aspect of a skill*".<sup>40</sup> La segunda de acuerdo a John Oller, "*is any procedure that causes the learner to process sequences of elements in a language that conform to the normal contextual constraints of that language, and which requires the learner to relate sequences of linguistic elements via pragmatic mappings to*

<sup>38</sup> Chomsky en Morrow, *Op Cit.*, p 150

<sup>39</sup> Baker, *Op Cit.*, p. 1.

<sup>40</sup> Carroll en John Oller, "Discrete Point, Integrative and Pragmatic Tests", p.36

*extralinguistic context* <sup>41</sup> En esta definición existen dos elementos importantes que deben estar presentes en un examen integrativo : primero, la tarea a realizar debe promover que el candidato utilice los diferentes aspectos de la lengua (las palabras, los sonidos , la sintaxis) en conjunto con las libertades y limitaciones de la lengua en particular ; segundo, para llevar a cabo lo anterior, el candidato decidirá, de acuerdo a sus conocimientos no sólo de la lengua sino de otros aspectos extralingüísticos (como las convenciones sociales), la forma en la que elaborará su discurso.

Tiempo después Oller propuso en 1979, lo que él llamó la "*Unitary Competence Hypothesis*" o UCH,<sup>42</sup> a partir de la cual el conjunto de exámenes utilizados en el enfoque psicométrico podían ser reemplazados por un sólo examen que cubriera la evaluación de la competencia comunicativa como un todo. A este tipo de examen se le llamó integrativo o pragmático. Baker enumera las ventajas del examen integrativo sobre el psicométrico de la siguiente manera:

- Mientras que la competencia medida por los exámenes psicométricos era divisible, la competencia que propone Oller es unitaria.
- Mientras que el examen psicométrico elimina el contexto, en el examen integrativo tiene un papel principal
- Mientras que el examen psicométrico requiere de técnicas complejas y forzadas, el examen integrativo utiliza técnicas simples y directas como el *cloze* y el dictado <sup>43</sup>

Reiterando lo dicho anteriormente, para Oller el desempeño de las personas en la lengua extranjera depende de lo que llamó "*pragmatic expectancy grammar*" o un sistema psicológico y real que ordena los componentes lingüísticos secuencialmente en el tiempo y en relación al contexto extralingüístico de manera significativa para la persona.<sup>44</sup>

A pesar de la discusión establecida, Oller ofrece una reconciliación entre las posturas psicométrica e integrativa. Lo que las diferencia es la base teórica, pero en formato pueden ser

---

<sup>41</sup> *Id*

<sup>42</sup> Baker. *Op. Cit.*, p. 69.

<sup>43</sup> *Ibid.*, p. 70.

<sup>44</sup> *Ibid.*, p. 69.

bastante similares. Ambos tipos de examen no son diferentes para fines prácticos, pero las bases teóricas contrastan. Las predicciones que hacen en relación a los efectos y a la validez relativa de los diferentes procedimientos y formatos también difieren fundamentalmente dependiendo del enfoque seleccionado.<sup>45</sup> Entendemos por esto último que aunque en el tipo de ejercicios que proponen no sean muy diferentes entre sí, el valor que le demos a la interpretación es lo que será significativo de acuerdo al enfoque seleccionado. Por ejemplo, podríamos decir que un alumno sobresaliente en un examen cuyo formato es del tipo psicométrico tiene ciertos conocimientos, ya sea sobre el sistema formal de la lengua o en cuanto al uso de la misma en funciones específicas. Pero la evidencia obtenida no podría utilizarse para afirmar que el mismo alumno puede utilizar la lengua correcta y efectivamente en situaciones reales de comunicación.

### 3. Consideraciones teóricas respecto al examen Lower.

Para los fines de la evaluación del inglés como lengua extranjera en la UAA concluiremos que el examen que se buscó diseñar es, en cuanto a su uso, un examen de aprovechamiento y al mismo tiempo un examen de dominio. El primero porque es un examen final que se aplica al término de los cursos para evaluar los conocimientos y habilidades aprendidos. Al ser un examen de aprovechamiento está basado en los contenidos del programa, por lo que su interés principal es un análisis de lo sucedido en los cursos. De acuerdo a esto hicimos entonces la selección de funciones y estrategias a evaluar para cada una de las habilidades. Sin embargo es también un examen de dominio para aquellos alumnos que no cursaron inglés en el CLE, y que por tanto no conocen el Programa.

De esta forma, el objetivo quedó definido de la siguiente manera. "Diseñar un examen de aprovechamiento que evalúe los conocimientos y habilidades de los alumnos en la lengua extranjera al nivel *Intermediate-mid*, que incluya las cuatro habilidades, gramática y vocabulario". Este examen aportaría la información necesaria sobre el desempeño de los alumnos en relación a los objetivos de enseñanza incluidos en el Programa Lower. Las habilidades de lectura y escritura estarían orientadas hacia el área de profesionalización de los alumnos. Incluiría una sección de gramática porque creemos que el conocimiento de la gramática es una ayuda para el estudiante que

---

<sup>45</sup> Oller. *Op Cit*, p. 36

esta tratando de comprender el sistema formal de la lengua extranjera. Atendiendo a esta razón, se utilizó un formato como la selección múltiple porque es un medio económico para evaluar una gran cantidad de estructuras a nivel de reconocimiento. Sin embargo, tomando en cuenta que un examen de gramática, como el propuesto en el Lower, no cubre el conocimiento y el uso de la lengua a nivel integrativo, se incluyeron también otras tareas, como la secciones de producción oral y escrita en donde el alumno, además de utilizar la gramática correctamente, tendrá que utilizar la lengua para un propósito de comunicación de manera efectiva.

## II. Proceso de elaboración del examen.

### A. Introducción.

#### 1. Propósitos del examen.

El programa Lower se empezó a aplicar desde febrero de 1992, y para el semestre enero-junio de 1993 la primera generación de alumnos estaba cursando el cuarto nivel. Por lo tanto a partir de esta fecha se dio inicio al proceso de elaboración del examen cuyo propósito es certificar los conocimientos de inglés de los alumnos. Este propósito es más complejo de lo que parece a simple vista ya que obedece a varios motivos. La certificación cumple con una función académica, administrativa y social.<sup>1</sup> En el caso del examen Lower, estos propósitos son los siguientes :

#### Propósitos académicos.

- Obtener información sobre los conocimientos aprendidos y no aprendidos por los estudiantes, y de esta manera predecir la actuación del alumno en una situación real. Es decir, el examen Lower contestará la siguiente pregunta : ¿En qué medida los estudiantes se acercan al nivel descrito como *Intermediate-mid* ?
- Elaborar un instrumento que, como parte de la evaluación, aporte información sobre la efectividad y factibilidad del Programa en relación a los alumnos
- Obtener información sobre la enseñanza, los materiales, las actividades y de la evaluación, ya fuera parcial o final, propuesta en el Programa.
- Modificar alguno de los aspectos anteriores de acuerdo a la información obtenida.

Lo anterior es de suma importancia porque de esta manera puede analizarse todo el conjunto de elementos involucrados en un curso. En nuestro caso, esta información ha sido muy valiosa, especialmente después de la segunda y tercera aplicación a los alumnos. En la primera aplicación se tenían alumnos que habían sido colocados directamente en los niveles 3 y 4, y que por tanto, no

---

<sup>1</sup> Arthur Hughes, *Testing for Language Teachers*, p. 48.

habían cursado los niveles anteriores en el CLE. Los resultados de esta primera aplicación se discutirán en el siguiente capítulo, pero es importante resaltar la diferencia entre el primer grupo de alumnos y las aplicaciones subsecuentes, donde los alumnos fueron ya el producto de los cuatro niveles de lengua en el CLE.

**Propósitos administrativos.** Incluyen aspectos de organización y revalidación institucionales:

- Obtener una calificación final para el cuarto nivel de inglés.
- Adquirir una constancia que acredite ante la propia UAA y la Dirección de Revalidación de la UNAM que se ha cumplido con el requisito de haber aprobado un examen de lengua extranjera necesario para la titulación.<sup>2</sup>
- Ofrecer una alternativa de acreditación para aquellos alumnos cuyo dominio de la lengua fuera equivalente o superior al nivel Lower, de manera que pudieran presentar un examen que certificara sus conocimientos sin necesidad de cursar los niveles establecidos

**Propósitos sociales.** La escuela, siendo una institución social, incide en la sociedad en la cual está inmersa. Por lo tanto, cualquier investigación, modificación o iniciativa tendrá necesariamente efectos en la comunidad. En el caso que nos ocupa sus efectos se previeron de la siguiente manera:

- Se establecería un nivel mínimo de dominio para los egresados de la UAA que les permitiera desenvolverse en su ámbito profesional.
- Como consecuencia del punto anterior, se haría evidente la necesidad de mejorar el nivel de lengua de los profesionistas que ya trabajan en el área con el propósito de incrementar la calidad de servicios que se ofrecen en Acapulco.

**Propósitos institucionales.** La UAA como institución pionera de educación superior en el sector privado en el estado de Guerrero tiene los siguientes propósitos:

- Establecer un nivel mínimo interno que sirva de referencia para cursos posteriores en la UAA.

---

<sup>2</sup> En un principio, se pensó que el examen podría cumplir con este requisito. Sin embargo, no fue así, por lo que los alumnos de ambas Facultades tienen que acreditar el examen de comprensión de lectura del CELE de la UNAM para titularse.

- Dar a conocer el Diploma Lower como una certificación que garantice el dominio (al nivel que el diploma indica) por parte del profesionista que lo presenta, y que describa en forma general sus habilidades en las áreas propuestas en cuanto a la lengua extranjera se refiere.
- Otorgar un reconocimiento a aquellos alumnos que se hayan distinguido en la presentación del examen

Como podrá observarse, todos estos propósitos influyeron en las decisiones en cuanto al tipo de examen que se diseñó. Por la importancia de estos propósitos y de las decisiones que de ellos resultaron, nos abocamos al diseño del examen de manera cuidadosa procurando cumplir además con los aspectos abordados en el capítulo anterior.

## 2. Proceso de diseño del examen Lower.

El examen Lower se elaboró siguiendo el procedimiento sugerido por Arthur Hughes que consta de las siguientes etapas: enunciar el problema, proponer la solución, reunir información sobre otros exámenes y analizarlos, sugerir posibilidades y escribir las especificaciones del examen, definir los contenidos, las operaciones, y de acuerdo a los candidatos seleccionar los temas, criterios, formatos y tiempos.<sup>3</sup> Es de acuerdo a estas especificaciones que la presente sección pretende describir tanto el proceso como el diseño mismo.

Recordemos que, como se mencionó en un capítulo anterior, el objetivo del examen fue evaluar los conocimientos y habilidades al nivel *Intermediate-Mid* propuesto por la ACTFL<sup>4</sup> y el Programa mismo. El diseño a cargo de la Mtra. Ordaz y la autora se llevó a cabo entre enero y junio de 1993 e incluyó las siguientes etapas:

- Los trabajos comenzaron con la revisión de la metodología de la enseñanza de lenguas extranjeras en el CLE y de los objetivos del programa incluyendo los cuatro cursos. De acuerdo a lo anterior definimos las áreas que se incluirían en el examen: comprensión auditiva, gramática, vocabulario, comprensión de lectura, producción escrita y oral

---

<sup>3</sup> *Id*

<sup>4</sup> Margarita Ordaz, *Proyecto de reestructuración de la D.L.E*, p. p 4-5

- Con la información anterior, recurrimos a la asesoría de la Mtra. Alma Ortiz Provenzal, en ese entonces Jefe de la Sección de exámenes del CELE de la UNAM, quien hizo las siguientes recomendaciones: Primero, para decidir sobre los reactivos que podrían utilizarse en la sección de gramática del examen, sugirió que se buscaran por lo menos dos exámenes de un nivel equivalente, que tuvieran la característica de haber sido validados con anterioridad y que fueran confiables. Se determinó que el examen de colocación diseñado por el CELE y que se aplica en la UAA y el examen TEAL de la Universidad de Michigan reunían las características deseadas dado que el examen de colocación hace una revisión de aquellos puntos gramaticales incluidos en los cursos de inglés de la preparatoria según el plan de estudios de nivel medio superior de la UNAM y algunos otros más de los propios programas del CLE.

Por otra parte, al utilizar un examen validado como el TEAL, se buscó reducir al mínimo la posibilidad de incluir reactivos que técnicamente no estuvieran bien contruidos y que por lo tanto, representarían una dificultad seria para los alumnos, lo que nos aportaría resultados poco confiables. Se aplicaron dichos exámenes a los estudiantes que estuvieran cursando el cuarto nivel en ese momento con la ayuda de los profesores del curso. Finalmente analizamos los resultados y decidimos, en base al programa y a la dificultad de cada reactivo, cuáles de éstos podrían ser incluidos en el examen.

- **Investigación sobre formatos y diseño de las secciones restantes.** Iniciamos una investigación documental sobre los formatos para cada sección del examen y seleccionamos los que consideramos los más apropiados de acuerdo a los objetivos y las tareas a realizar. Diseñamos cada sección del examen, incluyendo las instrucciones y los reactivos por sección junto con las hojas de respuesta para las secciones de comprensión auditiva, vocabulario y comprensión de lectura.
- **Primer piloteo.** Procedimos a hacer el primer piloteo con los profesores del CLE. Por razones de tiempo, este primer piloteo fue hecho tomando a este grupo que no pertenecía a la población

a la que se aplicaría el examen. El propósito fue detectar problemas en el formato, las instrucciones y la dificultad de los ejercicios, es decir se obtuvo evidencia sobre la validez aparente del examen. De acuerdo a estos resultados, se hicieron las correcciones pertinentes.

- **Escalas de calificación.** Acordamos la escala numérica que se utilizaría en la calificación del examen y tomamos la decisión de la puntuación por sección. Se determinó que la puntuación mínima para acreditar el examen sería del 60% y para obtener el diploma Lower sería de 70%. Lo anterior con el objeto de asegurar que los estudiantes que obtuvieran el diploma tuvieran un nivel representativo de lo que la UAA y el CLE se habían propuesto con sus estudiantes. Los alumnos que obtuvieran menos de 70% en el examen Lower, podrían cubrir el requisito de titulación fijado por la universidad, pero no obtendrían el diploma. También acordamos que para aquellos estudiantes que hubieran cursado el Nivel 4, su promedio del curso tendría un valor del 60% sobre el total de la calificación; el 40% restante sería determinado por la puntuación que obtuvieran en el examen Lower. Esta decisión se tomó en consideración al concepto de evaluación del CLE: el examen es sólo un instrumento más entre los que cuenta el profesor para otorgar la acreditación.
- Por último elaboré un instructivo para el examinador con información sobre la organización, tiempos, procedimientos, instrucciones, calificación, etc.

## **B. Descripción del examen.**

Después de haber sido piloteado entre los profesores, el examen Lower quedó conformado de las siguientes secciones:

1. **Sección de comprensión auditiva.** El objetivo es evaluar la obtención de información específica y detallada del texto auditivo. El texto auditivo seleccionado es un texto que sigue el formato de entrevista en una estación de radio. La conversación entre el locutor y su invitada versa sobre

los detalles más importantes de la vida de James Dean. Para seleccionar el texto se tomaron en cuenta los siguientes criterios:

- Que el texto estuviera en velocidad lo más parecida a la normal.
- Que el formato auditivo fuera lo más parecido a una situación real.
- Que fuera del interés de los alumnos.
- Que el tema no fuera completamente desconocido para los alumnos, pero que si aportara información nueva
- Que la cinta y la grabación fueran de buena calidad.

El formato de evaluación es un cuestionario de preguntas abiertas. El estudiante tiene que proporcionar la información específica que se le pide. El procedimiento es el siguiente: Primeramente el examinador pide a los alumnos que lean las instrucciones y las preguntas en el Anexo B de su hoja de respuestas. Después, el examinador reproduce la cinta tres veces, dando un minuto entre cada una para que los estudiantes puedan tomar notas. Finalmente, el estudiante responde las preguntas directamente en su hoja de respuestas. El tiempo dedicado a esta sección es de 15 a 20 minutos. Cada respuesta correcta tiene el valor de un punto con un total de 10 puntos en esta sección.

Con el fin de calificar adecuadamente esta sección y evitar las ambigüedades, se diseñó una hoja de repuestas para los examinadores, donde se especifican cuáles son las respuestas correctas, cuáles son aceptables y cuáles no lo son. Cabe hacer mención, que esta hoja de respuestas surgió después de la segunda aplicación como una necesidad que detectaron los profesores que calificaron los exámenes en ese entonces. Al ser un cuestionario de preguntas abiertas, se presentaron varias respuestas que podrían ser consideradas correctas.

**2. Sección de gramática.** El objetivo de esta sección es evaluar el conocimiento de los alumnos (a nivel de identificación) de los puntos gramaticales contenidos en los cuatro cursos del programa. Los criterios de selección que se utilizaron para decidir los reactivos que conformarían la sección fueron:

- Aquellos reactivos cuyas estructuras y funciones estaban incluidas en el programa.
- Aquellos reactivos de ambos exámenes (examen de colocación y TEAL) que demostraron tener, a través del análisis de reactivos, un índice de dificultad adecuado, es decir que se ubicaron en un intervalo de .3 y .7 en cuanto al índice de dificultad.<sup>5</sup>

El formato elegido es de selección múltiple con 3 distractores y la respuesta correcta. El tiempo estimado es de 35 minutos, teniendo en cuenta que los estudiantes invierten un minuto en cada reactivo. El valor de cada respuesta correcta es de un punto, lo que da un total de 35 puntos a la sección

**3. Sección de vocabulario.** Inicialmente se seleccionaron 15 reactivos del examen TEAL como resultado de un primer análisis de reactivos. El vocabulario seleccionado en esta primera versión era de inglés general, e intentaba evaluar el conocimiento de este vocabulario por medio de completar la frase con las opciones dadas, o de escoger el sinónimo de la palabra subrayada. El formato fue de selección múltiple con tres distractores y la respuesta correcta y el tiempo estimado fue de 15 minutos. El valor de cada respuesta correcta fue de un punto, por lo que el total por sección fue 15 puntos

Sin embargo, después de la primera aplicación en junio de 1993, el equipo de trabajo decidió modificar enteramente la sección de vocabulario, ya que estos resultados estaban afectando los promedios de los alumnos e incidían en la validez de contenido al no estar explícitamente contemplado en los cursos. Con la cooperación de los profesores que estaban impartiendo el 4o. nivel, el equipo de trabajo elaboró una lista de vocabulario mínimo para cada una de las carreras, es decir, una lista para Derecho y otra para Contaduría, Administración e Informática. De esta lista se seleccionaron 15 términos de cada una, y los profesores elaboraron los reactivos de selección múltiple.

---

<sup>5</sup>J. B. Heaton. *Writing English Language Tests*, p. 172.

En análisis posteriores se ha visto que ésta fue la decisión correcta. El vocabulario seleccionado es utilizado en actividades de lectura y expresión oral, por lo cual los estudiantes mostraron una mayor identificación y comprensión del vocabulario de su área. De este modo se aseguró la validez de contenido de esta sección del examen.

**4. Comprensión de lectura.** El objetivo de esta sección es evaluar la capacidad de los alumnos para realizar una lectura detallada y crítica de un texto de su especialidad. Dado el carácter de los temas, los estudiantes de Contaduría y Derecho tienen textos diferentes para trabajar. Se eligieron dos textos para cada Facultad y cinco reactivos de selección múltiple para cada uno de ellos que evalúan estrategias tales como buscar información específica y habilidades como el uso de la inferencia. Se calculó que los estudiantes se llevarían 60 minutos para leer y contestar las preguntas de ambos textos, y se designaron 10 puntos para esta sección. En la selección de los textos se tomaron en cuenta los siguientes criterios

- El texto debía ser auténtico, es decir que hubiera sido escrito para otros fines que no fueran la enseñanza de lenguas extranjeras.
- Que el texto estuviera relacionado con la carrera de los estudiantes.
- Que el tema fuera del interés de los estudiantes.
- Que el tema no fuera completamente desconocido para los alumnos y que aportara información nueva
- Que el vocabulario especializado necesario para la comprensión del texto estuviera incluido en el glosario de términos que los profesores manejan en clase.
- Que la impresión del texto sufriera la menor alteración posible. En el caso específico del TEXTO I de Contaduría “Mexican Workers” no fue posible mantener en su totalidad el formato original del texto, por lo que parte de este artículo fue capturado en computadora. De la misma manera, el TEXTO I “Judgement No. 145...” de la Facultad de Derecho, se capturó en computadora en su totalidad.

**5. Sección de producción escrita.** El objetivo de esta sección es evaluar la capacidad de los estudiantes para producir textos escritos en forma de ensayos cortos sobre temas relacionados con su especialidad en el que demuestren una mínima organización de sus ideas. De acuerdo a

esto, se pide a los estudiantes que desarrollen un ensayo breve (200 palabras) sobre uno de los temas que se dan a escoger. Los temas están relacionados con los temas discutidos en la sección de comprensión de lectura. En esta sección el alumno tiene que comparar, discutir o exponer su punto de vista sobre las preguntas o afirmaciones con que se expresan los temas.

Los aspectos que se toman en cuenta para evaluar esta sección son los siguientes: organización, ortografía, gramática y contenido. A cada uno de estos aspectos se le asignan 5 puntos, por lo que el total de puntos por sección es de 20 puntos.<sup>6</sup> Se calculó que los estudiantes utilizarían 30 minutos para escribir el ensayo.

6. **Producción oral.** El objetivo de esta sección es evaluar a los alumnos en el uso de la lengua extranjera oral de manera efectiva y concreta para narrar, describir, comentar, etc de manera informal sobre temas de interés general. Esta sección se lleva a cabo durante el examen en un área del salón, con un profesor responsable quien realizará las entrevistas. El procedimiento es el siguiente: El examinador explica en inglés al alumno lo que debe hacer, y mostrará al mismo una serie de fotografías que servirán de estímulo para que el alumno hable de 3 a 5 minutos sobre una o varias de ellas. El estudiante podrá describir, comentar, recordar, sugerir, reaccionar, opinar o argumentar sobre las fotografías. El examinador utiliza una hoja de evaluación donde aparecen los aspectos a considerar. Estos son: pronunciación, precisión gramatical, habilidad comunicativa, vocabulario y relación al tema. El total de puntos por sección es de 10 puntos.

El proceso de diseño y primer piloteo con los profesores concluyó en junio de 1993, y en el mismo mes de junio se llevó a cabo la primera aplicación a los alumnos cuyos resultados y conclusiones se presentarán en el siguiente capítulo.

---

<sup>6</sup> Para fines de análisis de esta sección posteriormente se adoptó una tabla de evaluación distinta a la propuesta originalmente. Ver siguiente capítulo.

## C. Cuadernillo de examen, instructivos y anexos.

### 1. Cuadernillo de examen.

El examen está impreso en un cuadernillo que se entrega a los alumnos al iniciar el examen junto con los anexos. El cuadernillo consta de: instructivo general, instrucciones por sección, los reactivos de las secciones de gramática, vocabulario (Contaduría y Derecho), comprensión de lectura y producción escrita. En estas dos últimas secciones también se hace una distinción entre los reactivos para Contaduría y Derecho respectivamente.

### 2. Instructivos.

Debido a la importancia institucional que desde un principio se decidió darle al examen Lower, la aplicación del mismo sigue todo un procedimiento detallado que le da el carácter de formalidad de requisito institucional. Este procedimiento inicia con la inscripción de los estudiantes al examen y finaliza con la entrega de resultados a los mismos. Por ser un proceso complejo, es necesario que el examinador tenga una idea clara de los pasos a seguir. Por otra parte también es de suma importancia que el alumno se entere del procedimiento del examen y algunas otras cuestiones del examen mismo. Para estos propósitos se diseñaron los instructivos que a continuación se detallan.

- **Instructivo para el examinador.** Este instructivo tiene por objeto el proporcionar al examinador la información necesaria para la aplicación del examen a los estudiantes y contiene la siguiente información: procedimiento de recepción de los alumnos en el aula asignada, procedimiento para la aplicación del examen, sugerencias para las instrucciones verbales y escritas del examen, requisitos administrativos, procedimiento e instrucciones para evaluar la producción oral de los alumnos y finalmente sugerencias a algunas posibles situaciones imprevistas durante el examen.

Este instructivo es revisado y comentado cada semestre en una reunión previa al examen en la cual los profesores que actuarán como sinodales revisan y actualizan los conceptos e información incluida en los instructivos anteriores.

■ **Instructivo para el alumno.** El objetivo de este instructivo es el de proporcionar al alumno información general sobre el examen, cómo está dividido, con qué tiempo cuenta para cada sección, el procedimiento a seguir, y algunas sugerencias útiles. Al inicio del examen, el sinodal responsable lee este instructivo junto con los alumnos y aclara las dudas al respecto. Este es el único momento en el que alumno puede preguntar acerca de cuestiones generales del examen. En todo momento, la lengua utilizada por los profesores y los alumnos es el inglés. El instructivo que aparece al inicio de cada cuadernillo de examen también está escrito en la lengua extranjera.

■ **Instructivo administrativo.** El propósito de este instructivo es el de proporcionar a profesores, estudiantes y comunidad universitaria en general la información necesaria para llevar a cabo la inscripción y registro al examen. La información incluye: *fechas, horarios y procedimiento de aplicación, requisitos, inscripción, registro y costo del examen.* También se proporciona información sobre lo que debe hacer el alumno el día del examen, dónde debe presentarse, qué debe llevar consigo, y finalmente la fecha de publicación de resultados finales. Por la variación de los datos mencionados, este instructivo es revisado y actualizado semestralmente.

### 3. Anexos.

Los anexos del examen Lower son aquellos documentos que no vienen incluidos en el cuadernillo de examen que se entrega a los alumnos. El propósito de dichos anexos es servir como hoja de respuestas de las diferentes secciones del examen y son los siguientes:

- **Anexo A.** Este anexo es la hoja de respuestas de las secciones de selección múltiple :gramática, vocabulario y *comprensión de lectura.*
- **Anexo B.** Este anexo contiene las instrucciones y preguntas de la sección de comprensión auditiva.
- **Anexo C.** Consta de dos hojas en blanco con el sello de la Universidad mismas que se utilizan para contestar la sección de producción escrita

#### **D. Recursos humanos, materiales y financieros.**

Dentro de la planeación del examen, es muy importante tomar en cuenta los recursos con los que cuenta el diseñador y la institución para la planeación y la aplicación del mismo. De alguna manera, la viabilidad de los recursos afecta de manera significativa el proceso, y determina algunas de las decisiones sobre procedimientos, tiempos, instalaciones, examinadores disponibles, equipo electrónico, fotocopias, etc

En relación al examen Lower, las autoridades sugirieron que el examen se mandara imprimir y encuadernar, lo que resultó muy conveniente dada la complejidad y extensión del examen; en segundo lugar, se acomodaron sólo 10 alumnos por salón para asegurar el control por parte de los profesores sinodales y para permitir que el examinador que llevara a cabo las entrevistas examinara sólo a pocos alumnos; como consecuencia de lo anterior los profesores sinodales tendrían pocos exámenes que calificar lo que se traduciría en calificaciones más confiables. De esta manera los recursos necesarios para la aplicación y calificación del examen son los siguientes:

- **Recursos materiales:** salones (un aula por cada 10 alumnos), cuadernillos de examen (uno por cada alumno), hojas de respuesta y anexos para el alumno y el profesor, instructivo para el profesor, copias de las listas de alumnos que tomarán el examen en esa aula, claves de calificación o plantillas para los reactivos de selección múltiple, criterios para calificar la sección de comprensión auditiva, fotografías para la sección de producción oral, hoja de evaluación de producción oral, grabadoras (una por salón) y cassettes con el texto para la sección de comprensión auditiva.
- **Recursos humanos:** profesores sinodales (dos por cada salón de 10 alumnos), un responsable de la organización (la Secretaría académica) y una secretaria.
- **Recursos financieros:** honorarios de los profesores sinodales (en el caso de que no sean profesores de tiempo completo), adquisición y/o reposición del material.

### III. Análisis de las tres primeras aplicaciones del examen Lower.

#### A . Introducción.

El propósito del capítulo que inicia es presentar los resultados obtenidos en el análisis estadístico de las tres primeras aplicaciones del examen Lower a los alumnos de la Facultad de Contaduría, Administración e Informática y de la Facultad de Derecho, en el periodo comprendido de junio de 1993 a junio de 1994. Estos resultados nos permitieron establecer la confiabilidad del examen como instrumento de evaluación y la validez del mismo en relación a los objetivos del curso. De la misma manera, haremos algunos comentarios y sugerencias surgidos de estos resultados con el fin de hacer este instrumento lo más confiable posible, en cuanto a sus componentes se refiere, en futuras aplicaciones. Para cumplir con este propósito se propone el siguiente orden :

- 1 Condiciones de aplicación y descripción de las poblaciones.
- 2 Reporte de resultados totales y por sección, incluyendo la media aritmética.
- 3 Conclusiones y recomendaciones.

Es importante mencionar que el análisis se hará en forma comparativa de las tres aplicaciones, estableciendo similitudes y diferencias. En cuanto al análisis estadístico se refiere, se tomó como referencia el trabajo de J. B. Heaton, *Writing English Language Tests*, de acuerdo al procedimiento sugerido en el Capítulo X y, excepción hecha al índice de discriminación, a los valores de referencia propuestos por él mismo para el análisis de reactivos. Estos son :

- **Media (x)**, cuyo resultado es el promedio aritmético de los resultados obtenidos por todos los candidatos.
- **Moda**, se refiere a la puntuación obtenida con más frecuencia por los candidatos.
- **Mediana**, se refiere a la puntuación obtenida por el candidato cuyo resultado lo ubica exactamente a la mitad del grupo en orden descendiente de calificación.

- **Desviación estándar (s.d.)**, mide el grado de dispersión de cada una de las puntuaciones en relación a la media. Esta medida es particularmente importante porque proporciona información sobre el grupo de estudiantes a los que se les aplicó el examen, como por ejemplo, si la medida es muy amplia estará describiendo un grupo heterogéneo cuyas competencias se encuentran dispersas y separadas unas de otras. En un examen de aprovechamiento, como el Lower, es deseable que la desviación estándar no sea tan amplia debido a que se presume que los candidatos han pasado por un proceso formal de enseñanza-aprendizaje en el salón de clase.
- **Rango**, indica la diferencia entre la puntuación más alta y la más baja. De la misma manera que la medida anterior, es deseable que la diferencia entre ambas puntuaciones no sea muy amplia, lo que indicaría una población muy heterogénea.

Por otra parte, para el análisis de reactivos se tomaron en cuenta los siguientes índices :

- **Índice de facilidad**, muestra el grado de dificultad o facilidad de un reactivo y expresa el porcentaje de estudiantes que contestaron el reactivo correctamente. Para nuestro propósito, tomamos como referencia el intervalo .3 a .7 como adecuado. Menos de .3 se consideró un reactivo difícil y más de .7 se consideró un reactivo fácil. Es importante mencionar que dicho intervalo se decide de manera arbitraria según el propósito del examen en cuestión:

*Although an average facility value of .5 or 50% may be desirable for many public achievement tests and for a few progress tests (depending on the purpose of which one is testing), the facility value of a large number of individual items will vary considerably. While aiming for test items with facility values falling between .4 and .6 many test constructors may be prepared in practice to accept items with facility values between .3 and .7<sup>1</sup>*

- **Índice de discriminación**, indica en qué medida un reactivo en particular separa o discrimina a los candidatos más sobresalientes de aquéllos menos sobresalientes. Convencionalmente, el índice de discriminación ideal es de 1.0, por lo que a medida que el resultado de la fórmula para determinar el índice se le acerque, se considera más adecuado al reflejar con más claridad la forma como se estructura la población motivo de estudio. Los reactivos que presentan un

índice de .3 o menos se considera que no discriminan en modo alguno, ya sea porque resultaron demasiado fáciles o difíciles para todos los candidatos en ambos grupos superior e inferior.

Sin embargo, en el presente trabajo fue necesario modificar este criterio por la razón expuesta a continuación. Tratándose de un examen de aprovechamiento cuyos contenidos fueron practicados en clase, pensamos que lo esperado es que la mayoría de los candidatos contesten correctamente los reactivos, lo que se expresa con un índice de discriminación cercano a 1.0. Por lo tanto, en el caso de las secciones de gramática, vocabulario y comprensión de lectura, en el reporte de resultados solamente recomendaremos revisar aquellos reactivos cuyo índice de dificultad haya sido muy alto. Asimismo, los reactivos cuyo índice de facilidad sea .7 o más (fácil para la población), y que el índice de discriminación sea .3 o menor, fue considerado adecuado

Los resultados se presentarán en el siguiente orden :

- Se hará un análisis de los resultados globales del examen cuyo propósito será determinar el funcionamiento del examen en general, para las poblaciones de la Facultad de Contaduría y para la Facultad de Derecho. Asimismo, se incluirá una gráfica que indique cómo quedaron distribuidos los exámenes con respecto a la media .
- Para las secciones de comprensión auditiva, gramática, vocabulario (excepto la 1ª aplicación) y comprensión de lectura, incluirán los resultados del análisis de reactivos con el fin de determinar la efectividad de cada uno de ellos.
- Para la sección de escritura, se presentarán los resultados de acuerdo al resultado de la clasificación propuesta en este trabajo. Cabe hacer mención que dicha decisión se tomó a partir de la imposibilidad de hacer un análisis y comentarios sobre los resultados obtenidos en esta sección utilizando la calificación numérica propuesta en un principio por el equipo de trabajo.<sup>2</sup> La tabla que aparece a continuación consta de tres niveles,<sup>3</sup> siendo el nivel 3 el máximo y el nivel 1 el mínimo. Para los propósitos de análisis, se consideró los niveles 2 y 3 como satisfactorios y el nivel 1 como insuficiente

---

*Ibid.* p. 173.

<sup>2</sup> En el Capítulo II de este trabajo se hace referencia a este problema de evaluación.

<sup>3</sup> Esta escala fue desarrollada como parte del proyecto *Bagrut Alternative Project, International English language Institute, Hunter College, City University of New York* 1986.

	<b>FLUENCY</b>	<b>ACCURACY</b>
<i>Nivel 1</i>	Vague statement of author's main intent (not necessarily explicit). Can deduce author's intent. No supporting information. No expansion of ideas. No awareness of organizational features (opening, body, wind-up). Lacking cohesion.	Poor grammar and vocabulary strongly interfering with comprehensibility. Elementary errors.
<i>Nivel 2</i>	Main intent expressed. Supporting ideas expressed. No expansion of ideas. Ideas not always clearly expressed. Lacking some organizational features. Mostly lacking cohesion	Frequent consistent errors yet comprehensible. Basic structures and simple vocabulary.
<i>Nivel 3</i>	Main intent and supporting ideas expressed clearly. Ideas generally expressed clearly. Occasional lack of cohesion. Exhibits proper organizational elements. Limited expansion of ideas.	Consistent errors. Accurate use of varied/richer vocabulary. Longer sentences structure.

Se incluyó además otra categoría , **Sin clasificar**, para todos aquellos trabajos en los cuales el alumno .

- No terminó la tarea.
- No entendió la tarea.
- Ambas.
- No respondió .

Para finalizar, se pensó que un análisis de frecuencia de selección de temas por los alumnos nos aportaría datos interesantes sobre las preferencias de los alumnos y tal vez, sobre el grado de dificultad de cada tema.

- Para la sección de producción oral se presentarán únicamente las puntuaciones obtenidas por los alumnos. No contamos con notas explicativas de los profesores ni tampoco con un seguimiento de la frecuencia de selección de temas.

## B. Descripción de las poblaciones.

### 1. Primera aplicación.

Como se ha mencionado anteriormente, la primera aplicación del examen se llevó a cabo en junio de 1993. En esa primera ocasión se presentaron 19 alumnos de la Facultad de Contaduría y 13 alumnos de la Facultad de Derecho, con un total de 32 estudiantes. Esta población presentó ciertas características que se mencionan a continuación:

- La edad de estos alumnos superó a la edad de las generaciones subsecuentes. Algunos de estos alumnos habían pasado directamente de la preparatoria a la Universidad, sin embargo un buen número de ellos habían interrumpido sus estudios por varios años.
- Los estudiantes de Contaduría estaban cursando el 3er. semestre de la carrera y los de Derecho el 4o. semestre.
- El conocimiento de la lengua que presentaron correspondía al 3°y 4° niveles en el examen de colocación aplicado por el CLE. Algunos de ellos habían estudiado inglés por diversas necesidades, muchas veces de tipo laboral; otros estudiantes habían viajado o vivido por un tiempo considerable en el extranjero. Un número reducido de ellos había estudiado los años correspondientes a la educación media superior, o uno o dos años de educación superior en los Estados Unidos o Canadá .

Debido a los resultados del examen de colocación, estos alumnos no cursaron los niveles 1 y 2 del Programa. En consecuencia presentaron el examen con menos horas cursadas de clases de comprensión de lectura y desde luego, sin el curso específico para esta habilidad. La versión del examen Lower que se les aplicó fue la primera versión piloteada entre los profesores a la que hicimos referencia en el capítulo anterior.

### 2. Segunda y tercera aplicaciones.

La segunda y tercera aplicaciones del examen Lower se llevaron a cabo en enero y junio de 1994. Para ambas aplicaciones tuvimos poblaciones similares de estudiantes :

- Las edades de estos alumnos también fue más alta que las de generaciones subsecuentes. De hecho, los alumnos de ambas aplicaciones pertenecen a la misma generación de las Facultades de Derecho y Contaduría. En este punto es importante mencionar que las tres poblaciones de estudiantes que analizaremos forman parte de una misma generación de ambas Facultades (1ª generación) dividida en 3 niveles por el examen de colocación.
- También se incluyen en estas aplicaciones algunos alumnos de nuevo ingreso a la Universidad.
- El conocimiento de la lengua que presentaron los alumnos de la 1ª. aplicación los colocó en el nivel 2 de inglés. De esta manera, al presentar el examen Lower habían cursado 300 hrs de inglés en la UAA correspondientes a los niveles 2, 3 y 4. Por otra parte la mayoría de los alumnos de la 3ª aplicación cursaron los 4 niveles del Programa Lower con un total de 400 hrs. de clase.
- Para ambas aplicaciones se utilizó la misma versión del examen anterior sin modificaciones con excepción de la sección de vocabulario.

En cuanto al número de estudiantes por Facultad, tenemos que se presentaron como sigue :

FACULTAD	1ª Aplicación	2ª Aplicación	3ª Aplicación	Total
Derecho	19	32	38	89 alumnos
Contaduría	13	24	14	51 alumnos

### C. Reporte de resultados.

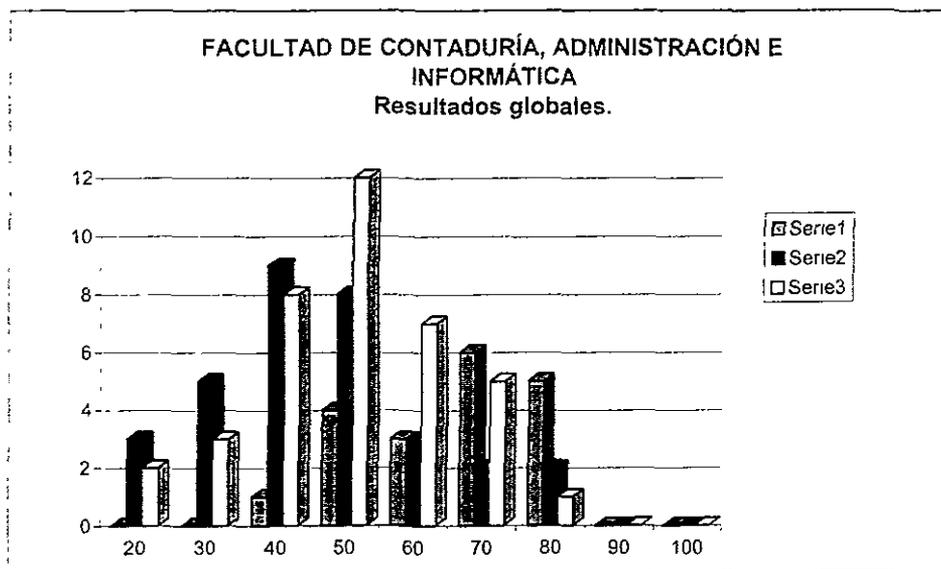
#### 1. Resultados globales.

#### FACULTAD DE CONTADURÍA

Examen	Media	Desv. Est.	Desv. Est. Corregida	Desv. Est. Corregida	Número
1ª Aplicación	69.47	86	71	11.97	37
2ª Aplicación	50.15	54.47	49	14.76	59
3ª Aplicación	46.66	60	55	16.07	58

Como se puede apreciar en el cuadro comparativo, el examen en su totalidad fue considerado fácil para la 1ª aplicación y más difícil para las dos aplicaciones posteriores, Aún así pareciera que los resultados son satisfactorios en la 2ª y 3ª aplicaciones ya que el 56.8% y el 50% aprobaron el examen respectivamente. Sin embargo, la media nos indica que, a pesar de que estos alumnos aprobaron, lo hicieron con puntuaciones bajas.

Por otra parte, la desviación estándar y el rango nos indican que las dos últimas poblaciones son más heterogéneas que la primera (ver gráfica siguiente), lo que sería de esperarse en un examen de colocación donde el desempeño de los alumnos puede ser muy diverso, pero no es lo deseable en un examen de aprovechamiento de un curso donde ha habido un trabajo de enseñanza aprendizaje previo y el examen representa los contenidos del sílabo.<sup>4</sup> Las observaciones anteriores están representadas en la siguiente gráfica:

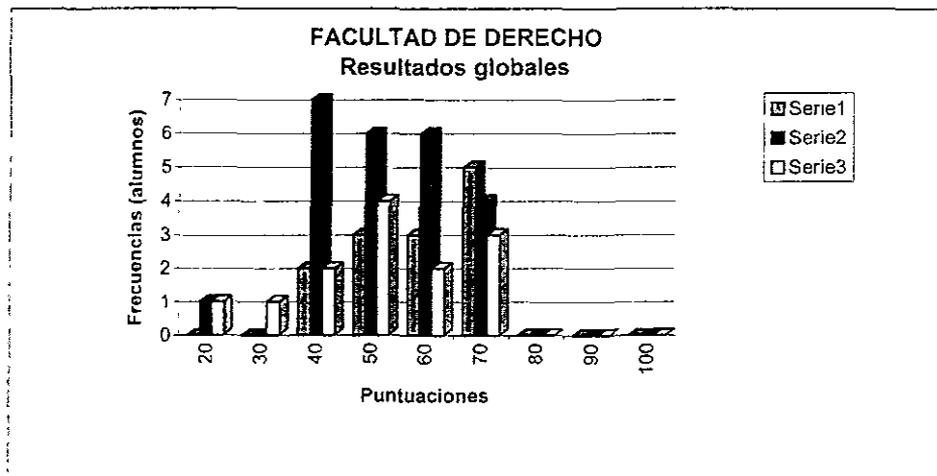


<sup>4</sup> Heaton. *Op Cu...* p

### FACULTAD DE DERECHO

	61.07	70	60	10.1	30
	55.75	60	55	57.27	54
	54.78	70, 53	53	9.7	52

En esta Facultad observamos que, de manera parecida a la Facultad de Contaduría, los alumnos de la 1ª. aplicación obtuvieron, en promedio, puntuaciones más altas que las dos aplicaciones subsecuentes, y entre éstas últimas existe poca diferencia entre ambas. Podría considerarse que el desempeño global de las dos poblaciones es satisfactorio, aunque los valores de la media indican que el promedio obtenido sigue dentro del intervalo no aprobatorio. Si tomamos en cuenta los índices de aprobación tenemos que 9 estudiantes (39.1%) y 5 estudiantes (35.7%) aprobaron en la 2ª. y 3ª. aplicaciones respectivamente. Particularmente llaman la atención la medida de desviación estándar para la segunda aplicación en donde vemos que los intervalos de separación entre un candidato y otro son sumamente amplios. Ambas medidas de dispersión, la desviación estándar y el rango, sugieren una población muy heterogénea cuyas competencias son también muy heterogéneas.



Los cuadros anteriores nos muestran que, de acuerdo al intervalo propuesto por Heaton, estos resultados se encuentran dentro del límite para considerar que el examen funcionó adecuadamente para ambas Facultades en cuanto a grado de dificultad se refiere. Sin embargo, vemos algunas diferencias. Por ejemplo, a pesar de que la Facultad de Contaduría obtuvo una media más alta en la 1ª aplicación (69.47) que la Facultad de Derecho para la misma aplicación (61.07). estos resultados no se mantienen en las aplicaciones subsecuentes, donde observamos que la Facultad de Derecho obtiene medidas de tendencia central más altas.

Por otra parte, las medidas de dispersión en esta última Facultad, desviación estándar y rango, nos indican que las puntuaciones obtenidas se agrupan de manera más uniforme alrededor de la media, y que la distancia entre el alumno que obtuvo la puntuación más alta y el que obtuvo la más baja es menor en la 1ª y 3ª aplicaciones, no así en la 2ª aplicación donde la desviación estándar sugiere una población sumamente heterogénea. Por lo tanto, se puede concluir que aunque los alumnos de la Facultad de Contaduría obtuvieron puntuaciones más altas, en general la población de la Facultad de Derecho resultó más homogénea.

Para poder ampliar la información anterior, es conveniente analizar cada sección del examen por sí sola, y finalmente en conjunto. Para realizar lo anterior, seguiremos el orden mismo del examen.

## 2. Resultados por sección.

### 2.1 Comprensión auditiva.

COMPRENSIÓN AUDITIVA		
	CONTADURÍA	DERECHO
	7.1 (71%)	6.2 (62%)
	6.53 (65.3%)	6.7 (67%)
	6.65 (66.5%)	6.4 (64%)

Como puede observarse, el desempeño de los alumnos de ambas facultades fue consistente. Con excepción de la 1ª aplicación de Contaduría, todos los demás obtuvieron medias

en un intervalo de 6.2 a 6.7, lo que en la propuesta de Heaton significa que en cuanto al índice de facilidad, esta sección resultó adecuada en cuanto a facilidad/dificultad se refiere.

En cuanto al análisis de reactivos, encontramos algunas coincidencias entre las tres aplicaciones para ambas Facultades :

- Los reactivos 3, 4, 5 y 10 parecen ser muy fáciles, por lo que su índice de discriminación indica que no discriminaron adecuadamente.
- El reactivo 7 sí discriminó adecuadamente
- El reactivo 8 fue considerado difícil en todas las aplicaciones. Esto pudiera deberse a que, en esta pregunta se pide a los alumnos que anoten los títulos de las películas que protagonizó James Dean. El problema es, que si el alumno no ha escuchado el título de las películas en inglés, difícilmente podrá identificar y escribir el título correctamente.

Las diferencias se presentaron en la forma siguiente :

- Se presentó una diferencia significativa entre la 2ª y la 3ª aplicaciones en relación al reactivo 1. Para la 2ª aplicación se obtuvo un índice de facilidad de .81. Sin embargo, para la 3ª aplicación obtuvo un índice de .34. Podemos observar que los resultados obtenidos se ubican en los límites del intervalo. En ambos casos el reactivo no discrimina.
- Los reactivos 2, 6 y 9 obtuvieron índices de discriminación muy bajos, considerándose difíciles para la mayoría de los estudiantes, grupo inferior y superior.

#### Recomendaciones :

- Modificar los reactivos 2, 6, 8 y 9.
- Modificar el reactivo 1 La diferencia entre los índices obtenidos en las diferentes aplicaciones parece ser debido a la variedad de posibles respuestas. A la pregunta : *Where is the program being held ?* La respuesta correcta es : *In KJIC Radio Station*. Sin embargo, también existen otras dos posibilidades según la clave de respuestas : *In a radio station*, o *In San Francisco* Esta variedad de posibilidades trae como consecuencia la confusión de los examinadores al

- respecto en detrimento de la confiabilidad, entendiéndose ésta como la independencia del resultado con respecto al examinador.

## 2.2 Gramática.

GRAMÁTICA		
	CONTADURÍA	DERECHO
	24.8 (70.8%)	23.07 (65.9%)
	16.1 (46.1%)	19.3 (55%)
	17.5 (50%)	15.2 (43.6%)

Como lo muestra la tabla anterior, tanto los alumnos de Contaduría como de Derecho obtuvieron la media más alta en la primera aplicación, que de acuerdo al intervalo sugerido, se considera adecuado con tendencia a ser considerado fácil. Vemos, sin embargo, un descenso en la media para ambas Facultades en las aplicaciones subsecuentes, lo que indica que para estos últimos alumnos el examen fue considerado más difícil. Dada la cantidad de reactivos en esta sección, el análisis de los mismos aportó datos muy diferentes para cada población. Especialmente, vemos una diferencia significativa entre la 1ª aplicación en comparación con las dos subsecuentes. Tomando en cuenta que estas dos poblaciones son más representativas la población de la UAA para la cual se diseñó el examen Lower<sup>5</sup> comentaremos sobre las similitudes y diferencias que encontramos en estas dos aplicaciones únicamente

En la Facultad de Contaduría encontramos las siguientes similitudes:

- Los reactivos 3 y 26 resultaron fáciles y por lo tanto no discriminan.
- Los reactivos 7, 10, 21 y 25 funcionaron adecuadamente en cuanto a dificultad se refiere y discriminaron de manera efectiva a los alumnos. Los reactivos 2, 6, 13, 22 y 27 obtuvieron un índice de dificultad adecuado, pero no discriminaron.

<sup>5</sup> Cfr. Capítulo III. Introducción

- Para la 2ª aplicación el reactivo 1 obtuvo un índice de discriminación de  $-0.06$  y para la 3ª aplicación obtuvo  $.05$ . Podríamos concluir que este reactivo no es confiable en cuanto a discriminación se refiere
- El reactivo 19 obtuvo en ambos casos un índice de discriminación 0. En situación similar se encuentra el reactivo 20 que obtuvo  $-.24$  y 0 respectivamente.
- Para ambas poblaciones hay una diferencia en cuanto a reactivos que pudieran considerarse difíciles.

En la 2ª y 3ª aplicaciones a los alumnos de la Facultad de Derecho encontramos las siguientes similitudes y diferencias :

- Los reactivos 23, 26, 30 y 35 resultaron ser fáciles y por lo tanto no discriminan. Especialmente los reactivos 23 y 26 obtuvieron índices de dificultad sumamente bajos.
- El reactivo 32 funcionó adecuadamente en cuanto a dificultad se refiere y sí discrimina. En cuanto al grado de dificultad, los reactivos 9, 13, 19 y 25 también funcionaron adecuadamente ; sin embargo en cuanto a discriminación se refiere, ninguno de ellos lo hizo de manera adecuada
- El reactivo 18 resultó difícil incluso para los alumnos más sobresalientes, por lo que no discriminó adecuadamente.
- Nuevamente los reactivos 19 y 20 presentaron problemas. El primero obtuvo 25 y  $-.14$  como índices de discriminación en la 2ª y 3ª aplicaciones respectivamente. El segundo obtuvo índices también muy bajos : 0 y  $.14$ . El reactivo 3 se encuentra en una situación similar al obtener como índices de discriminación  $-.27$  y 0.
- Las diferencias entre las medias de las dos aplicaciones, 54.6 y 43.28, son evidentes al comprobar que para la 3ª aplicación tenemos más reactivos considerados difíciles

#### **Recomendaciones para esta sección:**

- Mantener esta sección en su actual formato y versión, ya que parece funcionar de manera adecuada en cuanto a índice de facilidad se refiere.
- Modificar o reemplazar los reactivos 19 y 20.

- Modificar el reactivo 18 que parece ser muy difícil para los alumnos del grupo inferior y superior.

### 2.3 Vocabulario.

Debido a la modificación de esta sección, a la cual nos referimos en el capítulo anterior, únicamente se tomó en cuenta el análisis de la 2ª y 3ª aplicaciones para cada una de las Facultades. En la nueva versión se incluyeron términos propios de la especialidad de los alumnos, vocabulario que había sido visto y utilizado en clase. Fueron los profesores del CLE quienes reelaboraron esta sección de acuerdo a los siguientes criterios:

El objetivo de esta sección es evaluar el conocimiento de estudiantes sobre el vocabulario básico de su área de estudios. Deberá contener 15 reactivos de selección múltiple con valor de un punto cada uno en los cuales el estudiante demostraría su capacidad de identificación del vocabulario de su área de estudios. Algunos reactivos ofrecen una definición y el estudiante tiene que identificar cuál es el concepto definido. Otros reactivos proporcionan un ejemplo de uso del término en alguna oración, y el alumno tiene que identificar el sentido del término. Finalmente, algunos otros reactivos proporcionan ideas incompletas que el estudiante tiene que completar para que la frase tenga sentido y sea verdadera.

VOCABULARIO		
	CONTADURÍA	DERECHO
	6.4 (42.6%)	8.2 (55%)
	7.2 (47%)	10.6 (68%)

En general vemos que en la 3ª aplicación ambas Facultades presentaron un incremento en la media para esta sección. Esto podría deberse al hecho de que este vocabulario no es desconocido a los alumnos al ser incluido en sus clases en forma de material de lectura.

### Facultad de Contaduría, Administración e Informática.

Esta sección de vocabulario resultó ser más fácil para la 3ª aplicación que para la 2ª. Un dato interesante es que ambas aplicaciones resultaron muy similares, y como podemos apreciar, los resultados coinciden :

- El reactivo 1 resultó fácil para ambas aplicaciones.
- Los reactivos 6, 10, 11, 13 y 15 resultaron adecuados en cuanto a dificultad se refiere y el índice de discriminación muestra que casi todos los alumnos seleccionaron la respuesta correcta
- Según el análisis, el reactivo 7 presenta problemas. Para ambas poblaciones fue considerado muy difícil , situación que también se refleja en el índice de discriminación de 0 para ambas aplicaciones.

#### Recomendaciones para esta sección:

- Mantener esta sección en su estado actual. Realizar análisis de reactivos subsecuentes con el fin de determinar que el grado de dificultad se mantiene adecuado.
- Será conveniente revisar el reactivo 7 con especial cuidado y modificarlo si es necesario.

### Facultad de Derecho.

En esta sección los estudiantes de la Facultad demuestran reconocer bastante bien el vocabulario de su especialidad. Principalmente para los alumnos de la 3ª aplicación resultó más fácil este reconocimiento. Encontramos las siguientes similitudes :

- Los reactivos 8, 9, 10 y 13 resultaron fáciles para ambas poblaciones .
- El reactivo 14 resultó adecuado en cuanto al índice de facilidad se refiere.
- El reactivo 11 resultó ser muy difícil para ambos grupos superior e inferior.

#### Recomendaciones :

- Sera necesario observar esta sección en aplicaciones posteriores. Dada la gran cantidad de reactivos fáciles (no. 2, 4, 5, 8, 9, 10, 13) podemos suponer que el trabajo realizado en clase ha

sido muy efectivo, al revisar y utilizar este vocabulario en lecturas y tareas de escritura y expresión oral. Sin embargo creo necesario observar estos reactivos en detalle para asegurarnos que efectivamente el conocimiento del vocabulario por parte de los alumnos es determinante en el éxito de esta sección, y no las claves que pudieran obtener de los mismos reactivos y distractores.

- Solamente los reactivos 11 y 15 parecen ser difíciles para estas poblaciones. En el caso del reactivo 11, éste fue muy difícil tanto para los alumnos del grupo superior como del inferior.

#### 2.4 Comprensión de lectura.

Facultad de Contaduría, Administración e Informática.

COMPRESIÓN DE LECTURA		
	TEXTO I	TEXTO II
	2.1 (42%)	2.6 (52%)
	2.2 (43%)	1.6 (32%)
	1.4 (28%)	1.9 (39%)

El promedio de aciertos totales de los estudiantes de esta Facultad es de 47.5%. Este resultado refleja que para esta primera aplicación, las lecturas resultaron adecuadas al nivel de los alumnos. Para el TEXTO I titulado "The Mexican Worker" se tiene un índice de dificultad de 43%, mientras que para el TEXTO II "Special Controls to Safeguard Cash" el índice es de 52%. Esto indica que el segundo texto y los reactivos correspondientes al mismo resultaron comparativamente más fáciles que el primer texto. En la segunda aplicación tenemos que el índice de facilidad se mantuvo para el TEXTO I, no así para el TEXTO II, donde vemos un descenso significativo. Para la 3ª aplicación vemos que el TEXTO I resultó muy difícil para estos alumnos, y el TEXTO II podría considerarse adecuado.

En cuanto al análisis de reactivos, se obtuvieron los siguientes resultados :

- Para la 1ª aplicación, cada uno de los reactivos del TEXTO I presenta índices de facilidad entre un .4 y .5, con excepción del reactivo 4 que resultó ser muy difícil para los alumnos. En el TEXTO II, solamente el reactivo 1 resultó ser difícil para esta población.
- Los reactivos que fueron considerados adecuados para la 2ª aplicación, fueron considerados difíciles por la aplicación subsecuente. Tal es el caso de los reactivos 4 y 5 del TEXTO I. Del mismo texto, los reactivos 2 y 3 fueron considerados difíciles para las dos poblaciones.
- Al igual que en la 1ª aplicación, el reactivo 4 resultó ser muy difícil para las dos aplicaciones subsecuentes

#### Recomendaciones para esta sección:

- Del TEXTO II, revisar y modificar el reactivo 4 .
- Mantener esta sección en su versión y formato actuales.

#### Facultad de Derecho.

COMPRESIÓN DE LECTURA		
	TEXTO I	TEXTO II
	2.5 (51%)	1.7 (35%)
	2.5 (50%)	1.9 (38%)
	1.95 (39%)	1.6 (33%)

Como podemos ver, el TEXTO I "Judgement 145..." resultó adecuado para las tres poblaciones aunque es evidente que para la 3ª aplicación, el texto resultó más difícil. En contraste, el TEXTO II "Electoral System" parece mantener su índice de facilidad en un intervalo de .33 a .38, lo cual lo clasifica como adecuado. En cuanto a los reactivos, se obtuvieron los siguientes resultados en el análisis:

- Mientras que para la 1ª y 2ª aplicaciones el reactivo 1 del TEXTO I fue considerado fácil, para la 3ª fue considerado difícil.
- El reactivo 2 del mismo texto resultó muy difícil para las tres poblaciones
- Del TEXTO II, los reactivos 1 y 5 fueron difíciles para las dos últimas aplicaciones
- El reactivo 4 del mismo texto obtuvo un índice de discriminación 0.

**Recomendaciones :**

- Del TEXTO I , revisar y si es el caso, modificar el reactivo 2.
- Del TEXTO II, revisar y modificar el reactivo 4.

## **2.4 Producción escrita.**

Para el análisis y reporte de resultados de esta sección, se utilizó la tabla de evaluación propuesta al inicio de este capítulo. Sin embargo, es importante aclarar que para la 3ª aplicación la tabla de evaluación sugerida presentó algunas limitaciones. La más importante estuvo relacionada con la estructura de los trabajos de los alumnos. Es evidente, aún en los trabajos más deficientes, que los estudiantes tratan de seguir un modelo de organización producto de sus clases de escritura en inglés. El problema se presentó cuando la competencia lingüística, la puntuación y la ortografía de estos alumnos fue muy deficiente al punto de que el examinador no pudo entender el mensaje

Debido a lo anterior, los candidatos fueron clasificados en nivel 1. Sin embargo sí hubo un intento por organizar ideas, aún cuando no las expresaron coherentemente y con corrección. Para los fines de esta comparación con la 2ª aplicación, consideré conveniente no hacer ninguna modificación Sin embargo, en las conclusiones comentaré la necesidad de modificar la tabla de evaluación con el fin de adaptarla a nuestros estudiantes.

Facultad de Contaduría			
Producción escrita			
	1ª aplicación	2ª aplicación	3ª aplicación
	8 (42%)	4 (12.5%)	4 (10.5%)
	4 (21%)	12 (37.5%)	8 (21%)
	6 (31%)	5 (15.6%)	11 (28.9%)
	1 (5%)	11 (34.1%)	15 (39.4%)
	19	32	38

El cuadro anterior nos muestra que en cada aplicación se obtuvieron resultados diferentes. En el caso de la 1ª aplicación, tenemos que la mayoría de los alumnos se ubicaron en el Nivel 3, en contraste con las dos aplicaciones posteriores, en donde la mayoría se ubicó en los niveles 2 y 1 respectivamente. Especial mención merece el hecho de que en las dos últimas aplicaciones un porcentaje muy elevado de los candidatos no entendió la tarea o no tuvo tiempo suficiente para llevarla a cabo. De acuerdo a la tabla de evaluación propuesta, el trabajo de los alumnos se caracterizó de la siguiente manera:

- **Nivel 1.** En el caso de los estudiantes clasificados en este nivel, sus trabajos aún no reflejan el logro de los objetivos marcados por el programa. Su escritura presenta errores básicos de estructura sintáctica y ortografía. Algunos de ellos separan su escrito en párrafos visuales, aunque no sean ni cohesivos ni coherentes. La mayoría de las veces, se perciben los temas que interesan al alumno, aunque a veces no son expuestos claramente. En general, no existe una organización de ideas consciente por parte del estudiante, ni un desarrollo ni una conclusión a su trabajo.
- **Nivel 2.** Sus ensayos reflejan un intento de organizar sus ideas en forma lógica. Algunos de ellos tienen una introducción claramente definida, y en ésta logran exponer sus ideas de manera sencilla. Sin embargo, no hay una expansión de esas ideas, que se abandonan para introducir otras nuevas. Sus errores gramaticales y de puntuación hacen difícil la comprensión de algunas oraciones. Sus errores ortográficos son frecuentes, aunque es evidente que los alumnos conocen el vocabulario propio de su especialidad.

- Nivel 3. Estos alumnos presentaron pocos problemas en su producción escrita. Incluyeron una introducción, desarrollo y conclusión a su trabajo. La conclusión es claramente identificable. La mayoría de sus párrafos están terminados desde el punto de vista del desarrollo de ideas. Sin embargo, algunos de ellos quedan trancos. El uso del vocabulario de su especialidad es efectivo y exponen sus ideas claramente. En general, presentan pocos errores de ortografía, no así de puntuación, donde se ven algunos problemas. Utilizan, algunas veces con éxito y otras sin él, algunos conectores lógicos que introducen contraste, adición y conclusión.

Facultad de Contaduría			
Frecuencia de selección de temas			
	1ª aplicación	2ª aplicación	3ª aplicación
	7 (36%)	9 (28.1%)	20 (52.6%)
	3 (15%)	3 (9.3%)	1 (2.6%)
	1 (5%)	12 (37.5%)	8 (21%)
	3 (15%)	1 (3.1%)	4 (10.5%)
	5 (21%)	0	1 (2.6%)
	0	7 (21.8%)	4 (10.5%)

Como puede observarse, hay una marcada predilección por el Tema 1 en las tres aplicaciones. Esto podría deberse a que el Tratado de Libre Comercio o TLC es el punto a discutir durante los años 1993 y 1994. En el caso del tema 3, también favorito de los alumnos en la 2ª y 3ª aplicaciones, tenemos que el asunto a tratar es la situación de los trabajadores mexicanos en los estados Unidos. Esto podría explicar la predilección de los estudiantes por estos temas, que probablemente fueron objeto de discusión en algunas clases, incluyendo las de inglés. Sin embargo, es importante mencionar que particularmente en la 2ª aplicación, ningún alumno de los 12 que escogieron el tema se ajustó a las instrucciones. Todos ellos escribieron sobre las condiciones laborales en México, no en los EU. El Tema 2 consiste en exponer su punto de vista sobre la competitividad del mercado mexicano en relación al asiático, también tema muy discutido a raíz del TLC.

En el Tema 4 el alumno debe explicar y discutir las ventajas y desventajas de utilizar otros medios para salvaguardar el efectivo en los negocios. Posiblemente, este tema requiere de vocabulario técnico que los alumnos aún no conocían ni en inglés ni en español. Finalmente el Tema 5, en el cual deben mencionar y explicar los requerimientos para abrir una cuenta bancaria en algún banco mexicano, es un tema que los alumnos demostraron conocer bien, ya que incluso algunos de ellos hicieron la distinción entre diferentes tipos de cuentas bancarias

En cuanto a los errores gramaticales, de uso y de puntuación más comunes entre los estudiantes de esta Facultad están los siguientes :

- Los estudiantes tienen problemas en el uso de algunos modales como *must*, *will* y *may*. En el caso del primero, los estudiantes agregan la partícula *to* para las construcción *must +* infinitivo, p. ej. *"Mexicans must to study harder"* En los otros dos casos, hay una tendencia a utilizar ambos para expresar posibilidad en el futuro *"In the future maybe we will have better salaries."*
- Persiste el problema del lugar que ocupan los adjetivos en relación al sustantivo en inglés. Por ejemplo, *"Some people use the card credit (...) the majority have problems for pay<sup>6\*</sup> a percentage monthly."*
- Con marcada tendencia en la 3ª aplicación, tenemos que los estudiantes utilizan lenguaje cotidiano para expresarse en un ensayo académico que debería utilizar un registro más formal. Aunado a lo anterior tenemos el uso indiscriminado de contracciones: *"Many people consider that we aren't prepared for this trade and maybe they're right but if we don't look forward and try to demolish our problems we won't success and of coarse \*, we won't have a country with a good economy."*
- Algunos estudiantes, especialmente de la 3ª aplicación, podrían haber llegado al Nivel 3 de la tabla de evaluación. Sin embargo, quedaron clasificados en el Nivel 2 al no haber terminado el ensayo Por la forma abrupta en que parecen haber interrumpido su trabajo, concluimos que esto sucedió por falta de tiempo para terminar la tarea.

---

<sup>6</sup> El astensco (\*) indica que el error de ortografía aparece en el original

**Recomendaciones :**

- Es importante mencionar que un porcentaje significativo no terminó la tarea por falta de tiempo. Habrá que observar junto con los profesores sinodales si el tiempo estimado (30 minutos) es suficiente, y si además no hay alguna otra sección que esté tomando más tiempo que otras secciones.
- A consecuencia de lo anterior, habrá que promover en clase un mayor número de ejercicios de escritura controlados por tiempo.
- En el caso de los problemas gramaticales más comunes habrá que hacer énfasis en éstos en clase. Los problemas citados anteriormente están incluidos como objetivos en los programas.
- En cuanto a la selección de temas, habrá que observar las preferencias de las aplicaciones subsecuentes. El tema 1 parece estar sujeto al tiempo en cuanto a su importancia se refiere. Con esto queremos decir que el Tratado de Libre Comercio (TLC) era un tema controversial en las fechas de aplicación al estarse llevando a cabo las últimas negociaciones al respecto. Pudiera ser que debido a esto, el tema fuera muy significativo para los alumnos. Habrá que observar su funcionamiento en aplicaciones posteriores.

**Facultad de Derecho.**

<b>Facultad de Derecho</b>			
<b>Producción escrita</b>			
	1ª aplicación	2ª aplicación	3ª aplicación
	3 (23%)	1 (4%)	1 (7.6%)
	5 (38%)	6 (26%)	5 (38.4%)
	1 (5%)	5 (21%)	3 (23%)
	4 (30%)	11 (37%)	5 (38.4%)
	19	23	14

El trabajo de los alumnos de esta Facultad se caracterizó por lo siguiente :

- Nivel 1. Su trabajo muestra un intento por responder al tema, aunque responden como si fuera el tema una pregunta de la que se espera una respuesta breve. En ningún momento su producción escrita parece un ensayo, ni siquiera visualmente ya que no hay una separación visual de párrafos. Sus errores de ortografía y sintaxis son abundantes y demuestran el poco

dominio que tienen en estas tareas. Con frecuencia intentan traducir literalmente palabras o frases del español al inglés sin éxito. Exponen varias ideas diferentes sin relacionarlas o discutir las.

- Nivel 2. Casi todos presentan una introducción y desarrollo de sus ideas. Organizaron su escrito en párrafos visuales, aunque algunas veces esta separación es meramente visual. La mayoría de ellos completaron la tarea elegida y solamente parecen tener problemas en la conclusión, abrupta la mayoría de las veces. Tienen pocos errores de ortografía, aunque sí presentan errores en el uso de preposiciones, por ejemplo. La mayoría tiene problemas con la puntuación aunque esto no interfiere con la comunicación.
- Nivel 3. Estos alumnos llevaron a cabo la organización coherente de ideas, incluyendo una introducción, desarrollo y conclusión fácilmente identificables. Presentan errores de ortografía menores y algunos errores de uso. Demuestran el conocimiento de vocabulario de la especialidad aunque a veces no usan la palabra con la función gramatical que corresponde, por ejemplo "*politic*" en lugar de "*political*" o "*politician*". Estos alumnos completaron la tarea en forma satisfactoria en cuanto a contenido se refiere.

De la misma manera que los estudiantes de Contaduría tenemos a los alumnos que no terminaron la tarea, quizá por falta de tiempo. Sus ensayos se muestran interrumpidos a la mitad de una frase o palabra. Algunos demuestran no haber entendido la tarea. Otros intentan hacer un resumen de los textos presentados en la sección de comprensión de lectura, y algún otro escribe sobre un tema no incluido en la lista de temas de las instrucciones.

Además de los problemas citados para la Facultad de Contaduría, los problemas más comunes de los alumnos de esta Facultad en estas dos aplicaciones se encuentran los siguientes :

- Errores básicos de ortografía, p. ej. "*Cuauhtemoc Cárdenas and the candidate of the Pan are the must \*stronger*" en donde es evidente también una confusión de significado con la forma correcta

- Problemas con modales en la estructura *modal* + infinitivo. Por ejemplo, "*The citizens can to vote* "
- Falta de vocabulario de la especialidad como ilustra el ejemplo siguiente "*electoral boxes*" en lugar de "*ballots*".
- Uso de mayúsculas, por ejemplo : "*my opinion about Zedillo ponce de León is a Person that need more Practice in the politician.* "
- Confusión de los objetivos demostrativos "*this*" y "*these*" : "*... this parties in special...* "
- La puntuación erróneamente utilizada (o la ausencia de ésta) a veces dificulta la comprensión de la idea : "*And this is exactly our point. Because the electoral system is the way we have to become this participation reality.* "

<b>Facultad Derecho</b>			
<b>Frecuencia de selección de temas</b>			
	1ª aplicación	2ª aplicación	3ª aplicación
	7 (36%)	1 (4.3%)	0
	3 (15%)	0	1 (7.1%)
	1 (5%)	12 (52.1%)	10 (71.4%)
	3 (15%)	2 (8.6%)	0
	0	8 (34.7%)	3 (21.4%)

El Tema 1 consiste en explicar el proceso legal que un trabajador en México podría seguir si estuviera en el mismo caso que se describe en el Texto 1 "Judgement 145 ...". El Tema 2 consiste en explicar bajo qué circunstancias se rescindiría un contrato a un trabajador en Mexico. Es posible que la necesidad de vocabulario específico del área de Derecho Laboral fuera el problema por el cual un número muy reducido de candidatos decidió escribir sobre este tema. En contraste, los alumnos parecen conocer bien el tema 3, que consiste en explicar cómo son elegidos los candidatos a puestos públicos en México. Finalmente, el tema 4 pide al alumno comparar, mencionando similitudes y diferencias, ventajas y desventajas, los sistemas electorales ingles y mexicano. Es posible que este tema sea el más difícil de los propuestos porque pide comparar dos sistemas muy complejos, y que además el alumno debe conocer bien. El texto

"Electoral System" les provee la información requerida, pero no a profundidad. El propósito mismo del ensayo, es decir comparar, no es una función sencilla de llevar a cabo.

Se puede observar nuevamente el alto porcentaje de alumnos **sin respuesta**, especialmente en las dos últimas aplicaciones. Estos alumnos presentaron los mismos problemas que sus compañeros en Contaduría.

#### Recomendaciones :

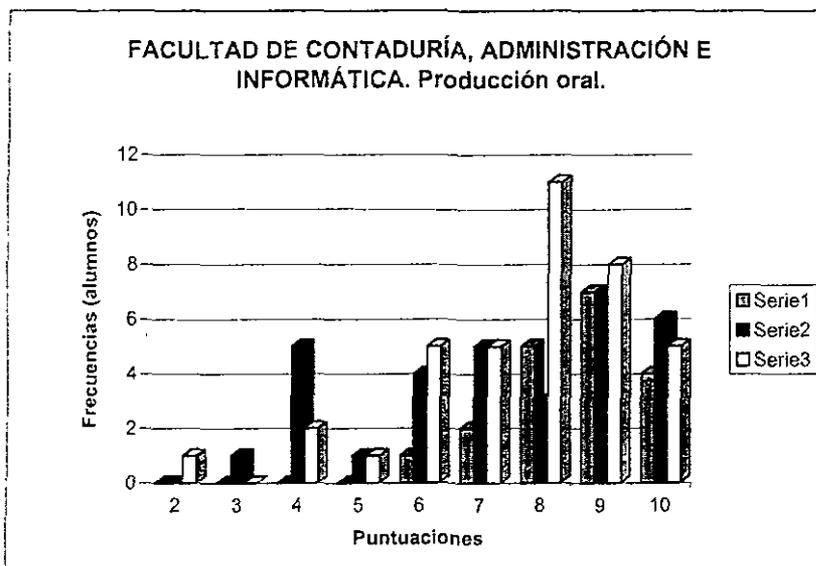
- Incluir en las clases de inglés mayor práctica de las estructuras y funciones que presentaron problemas para los alumnos. Algunas de ellas ya están contempladas en los programas, como el uso de los adjetivos demostrativos. De la misma manera, reciclar funciones y estructuras para aclarar las dudas de los estudiantes, y practicar aquéllas que son particularmente difíciles para éstos
- Es importante incluir ejercicios de puntuación en clase.
- El Tema 3 parece ser el favorito de los estudiantes. Creemos que es un tema rico y que los alumnos conocen bien. Habrá que observar en aplicaciones subsecuentes si este tema permanece en el interés de los mismos.

#### 2.5 Producción oral.

Facultad de Contaduría			
Producción oral			
	1ª aplicación	2ª aplicación	3ª aplicación
	4 (21%)	6 (18.7%)	5 (13.1%)
	7 (36.8%)	7 (21.8%)	8 (21%)
	5 (26.3%)	3 (9.3%)	11 (31.7%)
	2 (10%)	5 (15.6%)	5 (13.1%)
	1 (5.2%)	4 (12.5%)	5 (13.1%)
	0	1 (3.1%)	1 (2.6%)
	0	5 (15.6%)	2 (19.7%)
	0	1 (3.1%)	0
	0	0	1 (2.6%)
	<b>8.7</b>	<b>7.1</b>	<b>7.9</b>

Como se puede apreciar en la tabla anterior, la población de la 1ª aplicación es más homogénea al concentrar a sus candidatos en los tres primeros rangos, lo que no sucede en las otras dos aplicaciones en donde los candidatos quedaron repartidos de manera más heterogénea. Vemos también que, aún cuando existe esta diversidad en las aplicaciones 2ª y 3ª, un porcentaje alto de candidatos obtuvo la máxima puntuación. Las medias de cada aplicación indican que el grado de dificultad es el adecuado para las dos últimas aplicaciones. Para la 1ª, resultó ser muy fácil

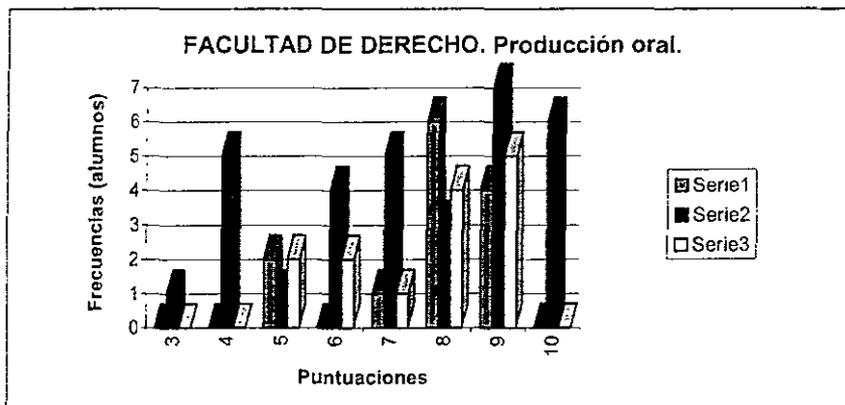
En contraste con la 1ª aplicación, tenemos un índice de reprobación del 21.8% y 24.9% respectivamente. Los resultados anteriores se ven reflejados en la gráfica siguiente :



## Facultad de Derecho

Facultad de Derecho Producción oral			
	1ª aplicación	2ª aplicación	3ª aplicación
	0	3 (13%)	0
	4 (30.7%)	6 (26%)	5 (35.6%)
	6 (46.1%)	2 (8.6%)	4 (28.5%)
	1 (7.6%)	7 (43.3%)	1 (7.1%)
	0	3 (13%)	2 (14.2%)
	2 (15.3)	1 (4.3%)	2 (14.2%)
	0	1 (4.3%)	0
	<b>7.8</b>	<b>8.1</b>	<b>7.7</b>

Como nos muestra la tabla, la 2ª aplicación obtuvo la media más alta, aunque también fue la población más heterogénea. Para las tres aplicaciones tenemos medias que nos indican que el grado de dificultad del examen en general es satisfactorio. El índice de reprobación es muy similar en la 1ª y 3ª aplicaciones, 15.3% y 14.2% respectivamente. Siguiendo la tendencia general, la 2ª aplicación presentó un índice menor: 8.6%.



#### IV. Conclusiones.

Habiendo expuesto los resultados del examen por sección, expondré a continuación las conclusiones y recomendaciones obtenidas del análisis estadístico con el fin de establecer el grado de confiabilidad y de validez del mismo como instrumento de evaluación y algunas reflexiones en relación al Programa Lower. Cabe mencionar que las conclusiones que presentaré a continuación son el resultado del análisis mencionado anteriormente y de mi experiencia directa en el Programa, tanto en el aspecto docente como en mis actividades como secretaria académica del CLE. Para apoyar las siguientes conclusiones, retomaremos los resultados de las tres aplicaciones

RESULTADOS POR SECCIÓN media por sección						
aplicaciones	CONTADURÍA			DERECHO		
	1ª	2ª	3ª	1ª	2ª	3ª
Comp. auditiva (10 pts.)	7.1 (71%)	6.5 (65%)	6.6 (66%)	6.2 (62%)	6.7 (67%)	6.4 (64%)
Gramática (35 pts.)	28.4 (70.8%)	16.1 (46.1%)	17.5 (50%)	23 (65.9%)	19.3 (55%)	15.2 (43.6%)
Vocabulario (15 pts.)	7.1 (70.8%)	6.4 (42.6%)	7.2 (47%)	6.7 (44.6%)	8.2 (55%)	10.6 (68%)
Comp. lectura (10 pts.)	5.2 (52%)	3.18 (31.8%)	3.5 (35%)	4.3 (43%)	4.3 (43%)	3.7 (37%)
Prod. Escrita (20 pts.)	15 (75%)	10.4 (52%)	11.7 (59%)	15 (75%)	8.8 (44%)	11.4 (57%)
Prod. Oral (10 pts.)	8.7 (87%)	7.1 (71%)	7.9 (79%)	7.8 (78%)	8.1 (81%)	7.7 (77%)
<b>Media global</b>	<b>69.47</b>	<b>50.15</b>	<b>46.66</b>	<b>61.07</b>	<b>55.75</b>	<b>54.78</b>

##### A. Acerca de los resultados globales.

Se observó en ambas Facultades un descenso significativo en la media de los resultados globales, descenso que se da de manera progresiva de una aplicación a otra. En este sentido podemos decir que el examen ha sido más difícil para las poblaciones subsecuentes. Desde el punto de vista del grado de dificultad, podríamos concluir que para los alumnos de la 1ª generación, los resultados del examen indican que fue considerado fácil, mientras que para la 2ª y

3ª aplicaciones estos resultados indican que el examen resultó con el índice de facilidad adecuado. Al obtener medidas de dispersión (desviación estándar y rango) más amplias es evidente que estas poblaciones (especialmente las de la Facultad de Contaduría) son sumamente heterogéneas, tal y como lo reflejan los resultados globales y por sección.

## **B. Acerca de los resultados por sección.**

### **1. Comprensión auditiva.**

Tomando en cuenta los resultados obtenidos, esta sección ha sido considerada como adecuada. Sería importante considerar el índice de facilidad en cada una de las preguntas y el alto número de alumnos que aprueban esta sección, que sería lo esperado tratándose de un examen de aprovechamiento. Sin embargo, será necesario asegurarse de que los alumnos están aprobando esta sección debido a que usan las estrategias y obtienen la información requerida del texto auditivo y no de otras fuentes. Por lo tanto, las sugerencias que me permito hacer a continuación están orientadas a establecer un mayor grado de confiabilidad a esta sección del examen, porque creo que es un problema importante en esta sección y que seguramente influyó en los resultados.

#### **Recomendaciones :**

- Cambiar totalmente el texto auditivo. Por una parte, los candidatos conocen ya el tema que se trata en el mismo, y algunos de ellos consultaron la biografía del personaje en cuestión y contestaron adecuadamente las preguntas. De esta manera, los candidatos ya venían preparados para contestar con información obtenida de otras fuentes, lo que nos proporcionó datos falsos sobre su habilidad para extraer información del texto auditivo. Por otra parte y para prevenir el mismo problema en el futuro, recomiendo tener por lo menos dos versiones del examen de comprensión auditiva. Pudiera tratarse del mismo texto con tareas diferentes o 2 textos diferentes con sus respectivas tareas.
- El formato elegido, es decir pregunta abierta, nos proporcionó algunos problemas : primero, la dificultad para distinguir las respuestas correctas de las que no lo son. Al elaborar la clave de

respuestas no se solucionó del todo este problema, como fue evidente al recalificar los exámenes. Algunos profesores siguen aplicando criterios distintos para calificar. Esto nos da origen al segundo problema el sistema de puntuación como consecuencia del formato elegido. Para el profesor que califica es sumamente difícil trabajar con puntos enteros y medios puntos, con respuestas correctas y parcialmente correctas.

Por lo tanto, mi recomendación es utilizar otros formatos de evaluación de la comprensión auditiva, tales como selección múltiple, llenar espacios en blanco u ordenar información; si se desca, incluir la pregunta abierta, pero no como formato único. Los formatos citados anteriormente reducen la posibilidad de utilizar información que no se encuentra en el texto y permiten la evaluación de diferentes estrategias, desde la identificación de información específica hasta la inferencia. Otras ventajas importantes son que reducen la subjetividad del evaluador y dan mayor oportunidad al alumno que pudiera tener problemas con ciertos tipos de formatos: "... we have to offer candidates as many 'fresh starts' as possible, and each task can represent a fresh start.(...) people will simply be better at some tasks than others."<sup>1</sup>

- Habrá que llevar a cabo un trabajo de academia con los profesores con el fin de discutir tanto las nuevas propuestas de examen como los aspectos que pudieran resultar conflictivos

## 2. Gramática.

Tomando en cuenta que en esta sección se incluyeron reactivos cuyas funciones están contempladas en los cursos, creo que el desempeño de los estudiantes no es el deseable, en cuanto a que el objetivo es que el alumno reconozca los exponentes presentados y revisados en clase. En la Facultad de Contaduría se presentó un descenso significativo de la 1ª a la 2ª aplicación, y una ligera recuperación en la 3ª. En la Facultad de Derecho, el descenso se presentó progresivamente de una aplicación a otra.

---

<sup>1</sup> Hughes *Op Cit.*, p. 81.

En cuanto al formato, la selección múltiple es una forma confiable, rápida y económica de incluir un mayor número de funciones en el examen, lo que nos asegura cierta validez de contenido. Podría sin embargo considerarse otros formatos para evaluar no sólo la gramática en una oración fuera de contexto, sino los conocimientos de los alumnos sobre la lengua extranjera para utilizar los diferentes componentes de la misma en el discurso. En este sentido sugiero incluir un ejercicio tipo *cloze* modificado, como lo propone Oller en *Language Tests at School* .

*In order to give correct responses (whether the standard of correctness is the exact word that originally appeared at a particular point, or any other word that fully fits the content of the passage), the learner must operate the basis of, both immediate and long-range constraints.<sup>2</sup>*

#### Recomendaciones :

Antes de hacer las recomendaciones pertinentes, cabe hacer mención que tuve la oportunidad de aplicar esta sección del examen a tres nativo hablantes norteamericanos que estuvieron de visita en la UAA. A pesar de las diferencias entre algunos críticos en relación a pilotear el examen con hablantes nativos, precisamente por la dificultad de definir el concepto "nativo- hablante", Charles Alderson *et. al.*, menciona que : "...for objective tests, native-speaker trials are invaluable."<sup>3</sup> Su propósito es evitar que en el examen hayan ambigüedades, que tanto el examen en sí como las instrucciones expresen lo que pretenden expresar.

Sin embargo, la selección de estos nativo hablantes ha de ser muy cuidadosa. En el caso que nos ocupa se buscó hablantes nativos de nivel educativo igual o superior al de nuestros propios alumnos. Seleccione finalmente a tres candidatos cuyas características fueron las siguientes : uno de ellos estaba cursando su tercer año en la universidad y los otros dos estaban en el primer año. Para esta prueba, apliqué el examen de gramática en el formato y versiones que se aplica a los alumnos en la UAA en el tiempo determinado para esta sección. Los tres jueces señalaron que encontraron problemas en los siguientes reactivos :

---

<sup>2</sup> Oller, *Language Tests at School*, p.42  
<sup>3</sup> Alderson *et al.*, *Op Cit.*, p. 97.

- En el reactivo 1 existe un problema de puntuación .

Dice "Jack's party was a lot of fun ! Really ! Who \_\_\_\_\_ ?"

De acuerdo a los jueces, el signo de admiración después de la palabra *Really !* debería cambiarse por un signo de interrogación. Al ser consultados al respecto dos profesores nativo hablantes, ambos estuvieron de acuerdo en que el uso de uno u otro signo estaba determinado por la intención del hablante. Por lo tanto no hice ninguna modificación al reactivo en este sentido. Sin embargo, si considero necesario hacer una distinción entre los participantes del diálogo con el fin de hacer más clara la intención y la participación de los interlocutores del intercambio verbal que se lleva a cabo. Por ejemplo,

- "Jack's party was a lot of fun !"

- "Really ! Who \_\_\_\_\_ ?"

- El reactivo 16 causó problemas a los jueces. En su opinión el reactivo tiene dos respuestas correctas :

*I have a lot of work. It's too important \_\_\_\_\_ any longer.*

- a) *that I leave it*
- b) *to leave them*
- c) *to leave it*
- d) *that I leave*

Según la clave de respuestas, la respuesta correcta es c. Los jueces apuntaron que la opción a también podría ser utilizada por un nativo hablante. Con el fin de aclarar la duda, consulté el libro *Practical English Usage* de Michael Swan.<sup>4</sup> De acuerdo a la fuente consultada, la respuesta correcta coincide con la clave de respuestas, es decir, c. Sin embargo, pudiera ser que un nativo hablante utilizara la opción a, especialmente en el lenguaje oral. Además está la ambigüedad de significado en el verbo *leave*. El problema podría resolverse cambiando este verbo por otro, por ejemplo *postpone*, pues si el alumno sabe que en la estructura *too + adjetivo* se usa un verbo en infinitivo, *to postpone it* sería la opción a elegir.

---

<sup>4</sup> Michael Swan, *Practical English Usage*, p. 263

Los alumnos examinados en las tres aplicaciones tuvieron problemas con el reactivo al considerarlo muy difícil. La opción a sí atrae a algunos candidatos, aunque no siempre a un número significativo de ellos. Por lo tanto, recomiendo cambiar la opción a o modificar el reactivo en su totalidad como sugerimos en el párrafo anterior.

- Se recomienda modificar los reactivos 19 y 20 de esta sección que de manera constante han demostrado tener dificultades en el diseño. Especialmente, la respuesta correcta para el reactivo 20 atrae a los candidatos del grupo inferior y no a los del grupo superior. En ambos reactivos, los otros distractores no están funcionando en modo alguno.
- Se recomienda tener por lo menos dos versiones diferentes de esta sección. Especialmente creo que sería útil hacer una investigación sobre la confiabilidad de esta sección por el método "split-half" sugerido por todos los autores consultados para la elaboración de este trabajo. De esta manera se aseguraría que ambas versiones tengan el mismo grado de confiabilidad.
- Considerar también la posibilidad de incluir otros formatos como el *cloze* o la identificación y corrección de errores.

### 3. Vocabulario.

De acuerdo a los resultados presentados, la sección de vocabulario parece estar funcionando en forma adecuada. Sin embargo, considero importante hacer las siguientes sugerencias:

#### Recomendaciones:

- Para la Facultad de Contaduría, el reactivo 7 presentó problemas en las dos aplicaciones. En ambos casos el reactivo resulta ser muy difícil. Veamos el reactivo y sus opciones:

*The innovative decision-making is feasible.*

- a *likely*
- b *successful*
- c *political*
- d *facilitate*.

De acuerdo a la clave de respuestas, la opción correcta es a. Sin embargo, el uso de la palabra *innovative* dirige a la mayoría de los estudiantes a la opción b. Recomiendo cambiar esta opción por alguna que no necesariamente se relacione con esta palabra (*innovative*), como por ejemplo, la opción c cuya función gramatical es posible pero cuyo sentido no parece ser adecuado. En este mismo reactivo la opción d no es posible, dado que la categoría gramatical es distinta a la palabra subrayada en la raíz.

- En esta sección encontramos algunos errores de captura. En el reactivo 2, la opción a dice *hundes* y debe decir *hunders*. En el reactivo 4, la raíz dice *Are* y debe decir *The*. En el reactivo 6, la opción b dice *control* y debe decir *to control*. Está de más decir que es muy importante corregir estos errores.
- Para la Facultad de Derecho, la mitad de los reactivos resultaron fáciles para ambas poblaciones. Sería conveniente realizar un estudio sobre estos reactivos para asegurar la validez de esta sección en cuanto a que el desempeño de los alumnos sea efectivamente debido al conocimiento del vocabulario de la especialidad. De manera especial habrá que revisar los reactivos 8, 9 y 10, en donde dos y hasta los tres distractores no funcionan como tales.
- En el caso del reactivo 10, la raíz dice : *The documentis were presented in court*. La palabra subrayada y el verbo nos indican que se trata de un sustantivo en plural. Sin embargo, las opciones que se ofrecen están en singular, por lo que ninguna es la correcta gramaticalmente hablando. Sugiero cambiar la palabra subrayada o modificar las opciones.

#### 4. Comprensión de lectura.

De acuerdo a los resultados estadísticos presentados en el capítulo anterior, podemos observar que en la 1ª aplicación los candidatos obtuvieron las puntuaciones más altas a pesar de no haber cursado el Nivel 1 orientado hacia la comprensión de lectura. Esta consideración es de suma importancia porque la hipótesis original sugería que por medio de la lectura en el nivel básico, los alumnos estarían en contacto con vocabulario y estructuras sintácticas más complejas,

que por otra parte no interferirían con la comprensión general de los textos. Esto traería como consecuencia a que el nivel de reconocimiento fuera mayor que su nivel de producción, por lo que en cursos posteriores estos alumnos demostrarían tener la capacidad para reconocer estas mismas estructuras. Al respecto, no se hizo ninguna investigación que comprobara si estos alumnos reconocían estas estructuras o si utilizaban sus estrategias de lectura correctamente, y cuáles de éstas últimas utilizaban en realidad. A lo que sí podemos referirnos es al logro de los objetivos propuestos en el Programa Lower.

En la Facultad de Contaduría tenemos que para la 3ª aplicación esta sección resultó más fácil. Sin embargo, la diferencia en la media de ambas aplicaciones es muy pequeña: para la 2ª tenemos una media de 3.18 y para la 3ª de 3.5. Esto indica que no hay una diferencia significativa entre los que llevaron el curso y los que no lo llevaron en cuanto a dificultad se refiere. Por otra parte, el porcentaje de alumnos que obtuvieron un mínimo 60% aprobatorio es en ambos casos muy bajo: el 12.5% y 10.5% respectivamente. De este 10.5% (4 alumnos) de la 3ª aplicación, tenemos que sólo uno de ellos cursó el Nivel 1. Los otros alumnos quedaron clasificados en los Niveles 3 y 4 del examen de colocación. En cuanto a los textos y sus reactivos se hacen las siguientes sugerencias:

#### Recomendaciones:

- Al revisar el TEXTO I "The Mexican Worker", encontré que hay un error de diseño en el reactivo 3:

3 *How would you compare productivity and quality between Mexico and the U. S. ?*

a) *equal*

b) *better*

c) *worse*

d) *underdeveloped*

El reactivo establece una comparación entre ambos países, pero esta comparación no está bien definida en las opciones b y c porque no establece cuál de los dos es mejor o peor en relación a los aspectos de productividad y calidad. Por otra parte la opción d no satisface la

comparación al carecer de la palabra que indica la comparación, es decir *more*. Sugiero modificar las opciones presentadas.

- En el TEXTO II “Special Controls To Safeguard Cash” el reactivo 4 resultó muy difícil para todos los candidatos. Al revisar el reactivo mencionado comprobé que su dificultad radica en que pone a prueba la capacidad del lector para establecer una relación de sinonimia entre las palabras *drawer*, *payor* y *depositor*. En el texto se explica quién es el llamado *drawer* pero no establece específicamente quién es el *depositor*. Mediante el análisis de cada uno de los sujetos involucrados en la transacción bancaria es que el candidato puede identificar la equivalencia de los tres términos. Tenemos entonces un reactivo que requiere utilizar la información dada para inferir una relación no explícita entre los tres términos, es decir, que este reactivo requiere de la utilización de la inferencia, habilidad propia de la lectura detallada y crítica

Al hacer un análisis extenso del reactivo, encontré que la respuesta correcta atrae a los candidatos del grupo inferior. Debido a esto último (y no al grado de dificultad) se recomienda revisar el reactivo y modificarlo.

- Sería conveniente utilizar otros procedimientos de captura o fotocopiado para el TEXTO I. Las gráficas en la primera página no aparecen claramente detalladas, lo que dificulta contestar la pregunta 3 ya mencionada anteriormente. Por otro lado, el cambio de un formato a otro en las páginas 8 y 9 dificulta la lectura a los alumnos, al tener que adaptarse al cambio sin previo aviso. Existen también algunos problemas de ortografía en la captura.
- Para la Facultad de Derecho tenemos resultados similares. Únicamente uno de los tres estudiantes (21.4%) que obtuvieron el 60% mínimo en la 3ª aplicación cursó el Nivel I. Retomando los objetivos del Programa Lower concluyo que siendo la lectura un proceso en el que intervienen tanto el conocimiento del mundo como del sistema formal de la lengua para obtener información, es evidente que en este caso el conocimiento de la lengua fue

determinante para el éxito o fracaso de la tarea, como fue evidente para los alumnos de la 3ª aplicación de ambas Facultades.

- Sería deseable un cambio en el tipo de letra, por ejemplo, para distinguir entre los textos y las preguntas para la Facultad de Derecho.

##### 5. Producción escrita.

En relación al logro de objetivos podría decir que de acuerdo a los resultados obtenidos estos son satisfactorios. Para la Facultad de Contaduría tenemos que un porcentaje de alumnos cumple satisfactoriamente con la tarea propuesta, aunque menos alumnos alcanzan el nivel 3 de la tabla de evaluación. Sin embargo, dada la complejidad de las tareas a realizar en el examen, el nivel 2 se ajusta al nivel de lengua que se pide. En la Facultad de Derecho se obtuvieron resultados similares, aunque es preciso decir que en comparación con los alumnos de Contaduría, menos alumnos alcanzaron el nivel 3.

De manera especial, me parece importante señalar el progresivo incremento en el número de alumnos que no entendió la tarea o no terminó el trabajo en ambas Facultades. En mi opinión, esto podría deberse a dos razones principalmente: la primera de ellas tiene que ver con la práctica de los alumnos para leer instrucciones, la segunda tiene que ver con el tiempo destinado para esta sección en relación a la complejidad de la tarea a realizar. Como comentaremos en recomendaciones siguientes, se requiere hacer un seguimiento al respecto con el fin de determinar el problema y plantear la solución apropiada al caso.

##### Recomendaciones:

Considero necesario revisar las instrucciones para esta sección con el objeto de verificar si las tareas están claramente explicadas. Igualmente sería deseable hacer énfasis a los profesores sobre el siguiente punto: probablemente los alumnos requieran más práctica para leer instrucciones correcta y eficientemente. El tiempo estimado por los diseñadores para esta sección es de 30 minutos. Sin embargo este límite de tiempo podría no ser el adecuado. Desde mi punto de vista, podrían existir dos problemas: que el tiempo fuera efectivamente inadecuado para la tarea a

realizar, o que los alumnos requieran más horas de práctica de esta habilidad medida con tiempo. Se recomienda estudiar si el tiempo sugerido es el adecuado para la tarea.

- Finalmente me parece importante hacer una recomendación en relación a la tarea misma propuesta en el examen especialmente para la Facultad de Contaduría (habida cuenta de que en la Facultad de Derecho el examen Lower ya no es un requisito para obtener la titulación). Esta propuesta va en relación a la hipótesis inicial para orientar el curso de escritura

En un principio se pensó que los estudiantes necesitarían la lengua extranjera para escribir ensayos académicos en relación a su especialidad. A través del tiempo y de algunas pláticas con profesores y alumnos, creemos que el curso de escritura pudiera orientarse hacia la realización de otras tareas tales como la elaboración de curriculum vitae, elaboración de reportes breves sobre actividades académicas o profesionales, redacción de algunos formatos de correspondencia comercial más usuales en el ámbito de los negocios, etc. Estas tareas sugeridas son de suma importancia para la vida académica y profesional de los alumnos de la Facultad de Contaduría. No descartamos la posibilidad de que los estudiantes de la Facultad de Derecho también se vean en la necesidad de redactar este tipo de documentos, además del ensayo académico.

En este mismo sentido enfrentamos un problema de validez y de confiabilidad. Aunado a la pertinencia del ensayo académico como único objetivo de los cursos de escritura, está el hecho de que incluimos una sola tarea, y en base a ésta, estamos prediciendo la capacidad de los candidatos para utilizar la lengua en una situación de comunicación, situación por demás cuestionable en el sentido de realidad. Sin embargo, lo que estamos evaluando es la capacidad del estudiante para redactar ensayos académicos. Con este tipo de tarea única podríamos estar evaluando con respecto a un criterio y no con respecto a la norma *"A norm-referenced measure will attempt to use a wide range of different items to sample a general field. (...) A criterion-referenced test, on the other hand, will not attempt to assess a comprehensive mastery of a field"*<sup>5</sup> Esta diferencia es de suma importancia porque habrá que definir si deseamos evaluar un

---

<sup>5</sup> Popham, *Op. Cit.*, p 32.

dominio específico o la habilidad de la escritura en general. Por cualquiera de las opciones que se opte, deberán especificarse los criterios y objetivos más claramente. Según el autor citado, la decisión depende del uso que se le de al examen y la información que se busca obtener. Al respecto sugiero estudiar una propuesta de cambio de los objetivos de enseñanza y por tanto de la evaluación.

- En cuanto a la calificación de los exámenes, mis recomendaciones se harán con el fin de conferir mayor confiabilidad a esta sección.

Primero, sería deseable contar con otro sistema de calificación para el examen. Se recomienda un sistema que, por una parte emita un juicio sobre el efecto producido por el trabajo como una unidad sin prestar particular atención a ningún aspecto en particular, escala como la que se utilizó para analizar los resultados en el presente trabajo, y por otra parte, una escala analítica donde otro examinador evalúe algunos componentes separadamente (gramática, ortografía, puntuación) en referencia a las descripciones que se ofrecen de cada uno de estos componentes.<sup>6</sup> Ambas escalas se pueden utilizar en conjunto para decidir sobre el desempeño de los alumnos. Esto último habrá de decidirse de acuerdo a los objetivos de la sección y a las decisiones involucradas.<sup>7</sup>

Asimismo será necesario adecuar las escalas a los objetivos fijados y a las características de la población en particular. Recordemos que uno de los problemas que enfrentamos en el análisis de resultados de la 3ª aplicación fue que la tabla de evaluación propuesta ya no era muy útil, al ser evidente que la competencia lingüística de los candidatos no estaba en relación directa a la organización con un propósito definido que presentaron en sus trabajos. Es notorio aquí el efecto producido por los cursos de escritura que, de acuerdo a los resultados obtenidos parece funcionar satisfactoriamente. Por lo tanto habrá que ser más específicos en cuanto a qué se espera de los alumnos en cada uno de los niveles en cuanto a la estructura del trabajo y a los otros aspectos que la academia de profesores considere relevantes.

<sup>6</sup> Alderson *et al.* *Op Cit.*, p. 108.

<sup>7</sup> Para propuestas sobre estas escalas, consultar Alderson, *Op Cit.*, y Jacobs, Holly *et al.*, *Testing ESL Composition: A Practical Approach* Newbury House Publishers, Rowley, 1980.

En segundo lugar, propongo que por lo menos dos sinodales revisen y emitan un juicio para cada uno de los trabajos. Esto no es difícil dado que para cada aplicación se manejan grupos de 10 alumnos y 2 profesores sinodales. Lo anterior disminuiría la subjetividad de un solo evaluador lo que redundará en puntuaciones más confiables. Considero también muy necesario mantener sesiones de entrenamiento y actualización del personal que califica los exámenes. Una actividad útil e interesante es fotocopiar algunos trabajos y discutir con los profesores sobre la calificación más apropiada en cada caso. Esto cumple con la doble función de actualizar al personal que ha aplicado el examen varias veces y entrenar al personal nuevo.<sup>8</sup>

**6. Producción oral.** Tomando en cuenta los resultados estadísticos obtenidos, es evidente que en todos los casos, en esta sección los alumnos obtuvieron las puntuaciones más altas. Es también la sección de la que menos información tuvimos ya que, aunque cada entrevistador tenía los criterios en su hoja de evaluación, éste hizo pocos o ningún comentario acerca de la producción de los alumnos. Creo que éste es un problema mayor porque no sabemos cómo han sido utilizados estos criterios, qué problemas han encontrado los profesores al aplicar el examen, etc.

Si analizamos las instrucciones que se le dan al examinador, vemos que éstas no especifican claramente la tarea: "El examinador explicará en inglés al estudiante lo que debe hacer y mostrará al mismo una serie de fotografías que servirán de estímulo para que éste hable 5 minutos sobre una o varias de ellas. El estudiante podrá: describir, comentar, sugerir, reaccionar, opinar, argumentar etc. sobre las fotografías libremente." Este tipo de actividad con infinitas posibilidades de interpretación y desempeño por parte del estudiante trae como consecuencia la poca confiabilidad en los resultados del examen. Por todo lo anterior, considero conveniente hacer algunos cambios con el fin de reducir la subjetividad, tanto del alumno como del entrevistador, y de la tarea misma a realizar.

#### **Recomendaciones :**

- En cuanto al diseño de la actividad creo que el establecer tareas más específicas ayudaría a alumnos y profesores a organizar sus ideas y comentarios más fácilmente. Para lo anterior,

---

<sup>8</sup> Alderson *et al.*, *Op. Cit.*, p. p. 110-113

podría utilizarse las fotografías ya propuestas pero con propósitos diferentes, como por ejemplo : comparar dos fotografías previamente escogidas por el examinador, describir en detalle alguna de las fotografías u ordenar secuencialmente algunas de ellas y explicar por qué se eligió ese orden.<sup>9</sup>

- En cuanto a la modalidad de entrevista de un sólo candidato con el entrevistador, podría modificarse a trabajar en pares. Esto nos brindaría otra oportunidad de interacción entre dos participantes y con ello, mayores posibilidades de utilizar la lengua para otras funciones como hacer preguntas y contestarlas, expresar acuerdo o desacuerdo, etc
- En cuanto a la calificación se refiere, ésta se mantendría de forma individual. Sin embargo, podría agregarse un aspecto más a evaluar como podría ser el logro o fracaso de la tarea a realizar cuando ésta dependa de la interacción de ambos alumnos. Por ejemplo, si éstos tienen que utilizar la información dada para ponerse de acuerdo o decidir sobre alguna situación en particular.
- Tanto la modificación de las tareas como de la dinámica nos proporciona diferentes opciones para obtener muestras más extensas y variadas. Esto evidentemente incidirá sobre el tiempo dedicado a la entrevista, que podría incrementarse.
- Ya sea que la entrevista sea individual o en pares, es conveniente hacer una propuesta de formato, es decir, en cuántas partes se divide la entrevista (en el caso de que se lleve a cabo más de una tarea), cuál es el objetivo de cada una y el tiempo estimado para cada una de ellas.
- En relación a los entrevistadores, la mayoría de los autores consultados proponen dos alternativas. La primera sugiere que haya 2 profesores examinadores. Uno de ellos entrevistará a los candidatos El otro sólo hará anotaciones y no intervendrá en la conversación. Al final de la entrevista, ambos examinadores cambian impresiones y llegan o no a un acuerdo. Otra variante es que el entrevistador evalúe el desempeño global y el observador evalúe los

---

<sup>9</sup> Nic Underhill, *Testing Spoken Language*, p 66

diferentes aspectos independientemente, tales como pronunciación, vocabulario, logro de la tarea y corrección gramatical, de manera similar a la propuesta para la sección de escritura.

La segunda opción propone que se graben las entrevistas y ambos examinadores otorguen la *calificación* posteriormente. La grabación de entrevistas resulta muy útil tanto en el entrenamiento y actualización de entrevistadores, como para propósitos de clase con los futuros candidatos al examen

- En cuanto a los criterios de evaluación podrían tomarse dos soluciones diferentes, y al mismo tiempo, complementarias. En primer lugar, la tabla que se maneja es muy útil porque presenta los aspectos a considerar en la evaluación. Sin embargo habría que definir claramente a qué equivale cada punto de la escala. Para complementar esta tabla, sugiero utilizar los diferentes niveles de descripción de las *ACTFL Proficiency Guidelines* citados anteriormente. Arthur Hughes<sup>10</sup> y J. B. Heaton<sup>11</sup> también proponen diferentes escalas en las obras citadas en este trabajo, algunas de ellas de desempeño global y otras analíticas, como las propuestas para la sección de producción escrita.

### C. Consideraciones finales.

Como me permití exponer en la introducción de este trabajo, la importancia y pertinencia del mismo radica en el análisis del examen como evaluación, no sólo del logro de los objetivos de aprendizaje en relación a los alumnos, sino de otros aspectos del proceso educativo, función que Popham resume de la manera siguiente: "*A test is merely a tool that we use in order to make inferences.*"<sup>12</sup> De esta manera, el examen nos aportó información útil sobre el Programa Lower, la efectividad de la enseñanza, la pertinencia de los materiales didácticos y por supuesto, el desempeño de los alumnos en cuanto a los objetivos de aprendizaje.

---

<sup>10</sup> Hughes, *Testing for Language Teachers*, p. 87-88.

<sup>11</sup> Heaton, *Op. Cit.*, p. 95

<sup>12</sup> Popham *Op. Cit.*, p. 94.

En este sentido, mis consideraciones finales estarán principalmente orientadas hacia dos aspectos :

1. Efectividad del examen Lower como instrumento de evaluación.
- 2 La efectividad del Programa Lower, es decir, si el Programa ayuda a cumplir la meta fijada : el uso de la lengua extranjera como una herramienta que el alumno pueda utilizar en el ejercicio de su profesión.

### 1. Respecto al examen Lower como instrumento de evaluación.

En este apartado me interesa concluir sobre los dos aspectos citados en el primer capítulo de este trabajo, y que son de vital importancia para determinar si el examen Lower ha funcionado para el propósito que fue diseñado: la confiabilidad y validez del examen como instrumento de evaluación. En este sentido, me refiero a Popham en la cita anterior, en donde especifica que el examen es sólo una herramienta para obtener evidencias y de esta manera, *hacer inferencias sobre el aspecto o aspectos que nos propusimos evaluar.*

Messick comenta al respecto que no se puede hablar de la validez de un examen, sino de la validez de las inferencias que hacemos a partir de la información obtenida : "*one validates, not a test, but an interpretation of data arising from an specific procedure*"<sup>13</sup> De esta manera queda entonces establecido nuevamente que la confiabilidad del instrumento es el punto de partida para poder obtener evidencias y por ende, una interpretación válida de la información.

En la sección anterior comentamos sobre las diferentes secciones del examen y me permití hacer algunas recomendaciones sobre cada una de ellas. Como aclaré en ese momento, las recomendaciones están orientadas a conferir mayor confiabilidad al examen. Efectivamente, aunque nuestro objetivo no fue diseñar un examen 100% objetivo, meta por demás inalcanzable desde nuestro punto de vista, sí buscamos que el examen nos proporcionara la evidencia de las habilidades y conocimientos que nos habíamos propuesto.

---

<sup>1</sup> Messick en Popham. *Op Cit.*, p 94

Al hacer un análisis estadístico del examen y sus diferentes secciones, algunas imprecisiones que podrían estar alterando los resultados fueron evidentes, y por tanto, la interpretación de estos resultados podría no ser válida del todo. Tal es el caso de las secciones de comprensión auditiva, producción escrita y oral. Aunque cada sección presenta problemas diferentes, todos ellos afectan de alguna manera a los resultados de las mismas. Por todo ello, es de vital importancia que tomen en cuenta las recomendaciones anteriores y otras que los mismos profesores y autoridades involucrados crean pertinentes.

De manera especial sugiero llevar a cabo otros estudios sobre la confiabilidad, utilizando para ello los procedimientos sugeridos por algunos autores. Entre éstos se encuentran investigaciones sobre la confiabilidad externa tales como la estabilidad del instrumento y el *proceso alternado de confiabilidad*; y de consistencia interna como el procedimiento de Kuder-Richardson en sus dos alternativas.<sup>14</sup> Como podemos ver, esta tesis se ha orientado hacia la consistencia interna del examen; sin embargo, sería deseable que el trabajo de investigación continuara en las otras direcciones mencionadas.

En cuanto a algunos aspectos de validez, podemos decir que las principales preocupaciones al respecto de los conceptos tradicionales de validez estuvieron orientados hacia la validez aparente, de contenido y de construcción. En este sentido se cumplió con la primera desde el momento del primer piloteo entre los profesores. En cuanto a la validez de contenido y de construcción, se buscó desde un principio que el examen fuera una muestra representativa de los contenidos de los cursos y que reflejara el enfoque de enseñanza de lengua extranjera adoptada por el CLE.

Además del trabajo anterior, creo que es necesario realizar un trabajo de investigación más sistemático, como el propuesto por Popham<sup>15</sup> y que aportaría información útil al respecto. Este trabajo se apoya en cuestionarios tanto para los candidatos como para los examinadores, y están orientados a obtener información sobre las habilidades y conocimientos

---

<sup>14</sup> Popham. *Op Cit.*, p. p 123-136.

<sup>15</sup> *Ibid* p.p 97-107.

que los candidatos hayan utilizado. Como ya mencionamos anteriormente, recomendamos especialmente que se lleve a cabo este tipo de análisis para la sección de vocabulario de la Facultad de Derecho. En el caso de los examinadores, la información que se pide es aquella relacionada con la representatividad del contenido de los cursos en el examen.

Al respecto de los dos tipos de validez mencionados en el párrafo anterior, me permito hacer algunas reflexiones. Reiterando lo dicho anteriormente, se buscó desde un principio ser congruente con la metodología y contenidos del Programa. Por ejemplo, en la sección de comprensión de lectura se prefirieron los textos auténticos sobre la especialidad de los alumnos a los diseñados para propósitos de enseñanza de lenguas o los diseñados para exámenes concebidos para otros propósitos.

Podría pensarse, *sin embargo*, que otras secciones, como la de gramática, no cumple totalmente con estas condiciones dado que el Programa Lower está organizado en funciones y nociones. Creemos, *sin embargo*, que la gramática está incluida de todas formas en un programa organizado en funciones y nociones y que es necesario entender el sistema formal de la lengua

extranjera para su aprendizaje y uso. Como se aclaró en su momento, la decisión de incluir este aspecto con un formato que recuerda los enfoques estructuralistas fue una decisión basada en la economía de medios para asegurar la confiabilidad, y hasta cierto punto la validez de contenido.

Por otro lado, también es cierto que el hecho de que un alumno identifique la respuesta correcta en esta sección del examen no permite predecir si este mismo alumno las utilizará de manera correcta y efectiva en una situación real de comunicación. Este comentario lo podemos hacer extensivo a la sección de vocabulario. La validez de la interpretación que se haga de estos resultados es la que nos preocupa. Concluyo entonces que en esta sección, el examen mide la habilidad de reconocer las estructuras gramaticalmente correctas, pero esto no quiere decir que el alumno posee necesariamente la competencia comunicativa de la lengua.

Para lograr un balance entre ambas posturas, creo que las secciones de producción escrita y oral pueden contrarrestar la evaluación analítica. En *Testing ESL Composition*<sup>16</sup>, los autores mencionan una lista de ventajas por las cuales conviene incluir una sección de escritura en un examen. Después de analizarlas pienso que son extensivas a la producción oral, ya que .

- Enfatizan el uso de la lengua extranjera para propósitos de comunicación sobre la producción de oraciones mecánicas y sin sentido.
- Promueven un acercamiento entre lo que se enseña y lo que se examina, es decir el uso de la lengua para propósitos de comunicación.
- Y finalmente es más válida que un examen por puntos específicos al proporcionar evidencia de la competencia comunicativa.

Oller agrega que la producción escrita "*[It is ] nearly certain to produce more meaningful and readily interpretable results.*"<sup>17</sup> Esta última aseveración es evidente en el análisis de los resultados de la sección de escritura. A partir de los resultados se pudo hacer un diagnóstico mucho más amplio y significativo de las muestras producidas por los candidatos, sus logros, sus errores más comunes y el funcionamiento general del programa de escritura.

Para concluir sobre este punto diremos que ambas secciones proporcionan este balance tan necesario en el examen, donde el grado de confiabilidad establecido por un formato de selección múltiple está en equilibrio con la validez que proporciona una actividad en la que ambas competencias lingüística y comunicativa se interrelacionan para producir el discurso oral o escrito teniendo como fin la comunicación.

Sin embargo aún queda un punto por resolver en esta tensión entre ambos conceptos. La sección de gramática tiene un valor de 35% sobre las otras dos, 20% de la sección de escritura y

---

<sup>16</sup> Holly L. Jacobs *et al.*, *Op Cit* ., p. 3.

<sup>17</sup> Oller. *Op Cit* , p. 227

10% de la de producción oral. El peso de esta sección y la de vocabulario<sup>18</sup> (15%) sobrepasa al peso otorgado a ambas habilidades productivas. Además en estas secciones, la corrección gramatical es uno de los aspectos incluidos en los criterios de evaluación. Propongo que se reconsidere nuevamente el peso de estas secciones en relación a su importancia dentro de la metodología y objetivos del Programa Lower y que en base a las investigaciones que se hagan al respecto determinar nuevamente la puntuación y el peso de estas secciones en el total.<sup>19</sup>

## 2. Respecto al Programa Lower.

Al iniciar la última sección de este trabajo, quiero reiterar que hablar de evaluar la efectividad del Programa Lower es un objetivo muy amplio y podríamos decir ambicioso, por lo que es necesario puntualizar que *me referiré solamente a las conclusiones a las que los resultados del examen me permitieron llegar.*

Desde el inicio del proyecto de la UAA como institución de enseñanza superior se fijó el objetivo de incluir los cursos de lengua extranjera como una ayuda a los estudiantes durante sus estudios y posteriormente en su vida profesional. Este objetivo marcó, por así decirlo, la orientación de los cursos de lengua extranjera que impartía la entonces Dirección de Lenguas Extranjeras, hacia cursos que fomentaran el uso de la lengua extranjera para la comunicación en situaciones de la vida cotidiana y del área de especialización de los estudiantes. Consecuentemente el proyecto contempló la necesidad de cursos globales en donde se incluyeran las cuatro habilidades.

Sin embargo, también el objetivo mismo indicó que era necesario reforzar algunos aspectos con propósitos específicos. Particularmente fue evidente la necesidad de los alumnos de consultar bibliografía especializada en inglés desde los primeros semestres. Por lo tanto, se decidió que paralelamente al curso global, se integrara un programa cuyo objetivo es la lectura de textos inicialmente de interés general y posteriormente del área de especialización de los estudiantes.

---

<sup>18</sup> *Id* El autor establece que los exámenes de vocabulario son también exámenes de gramática

Es por esta razón que el examen Lower incluyó textos de las áreas de Contaduría y Derecho y una sección de vocabulario específico de cada una. Por lo tanto, ha sido de especial interés para las autoridades de la propia Universidad y de los profesores del CLE el conocer los resultados del mismo para evaluar la efectividad de esta combinación *sui generis*. Sin embargo no es el objetivo del examen Lower proporcionar los *medios para investigar* sobre los efectos de haber incluido desde el inicio del curso la comprensión de lectura, la manera en que esta habilidad influya en el reconocimiento de estructuras presentadas en cursos posteriores así como tampoco el uso de estrategias de aprendizaje (*desarrolladas durante los cursos de comprensión de lectura*) significó para los alumnos que cursaron los cuatro niveles del Programa Lower.

Los estudiantes de la 1ª aplicación se desempeñaron mejor en general, a pesar de no haber cursado los dos primeros niveles. Estos alumnos tuvieron uno o dos semestres de práctica de esta habilidad dentro de los cursos globales de los niveles 3 y 4. Podemos decir que su nivel de lengua inicial fue factor determinante en su desempeño general en la lectura. Esta hipótesis es apoyada por los alumnos que aprobaron esta sección en la 3ª aplicación : con excepción de dos ellos, todos los demás iniciaron el curso en los niveles 3 y 4. Por lo tanto, concluyo que los resultados obtenidos por estos alumnos son consecuencia de su trabajo en los últimos dos cursos del Programa Lower y de su nivel inicial de lengua que les permitió obtener mayores beneficios de los niveles 3 y 4.

Por otra parte los alumnos de las aplicaciones subsecuentes muestran una competencia lingüística y comunicativa menos efectiva, afirmación que es sostenida por los resultados ya expuestos, donde el desempeño global y por sección es menor para estas dos aplicaciones. Considerando que el programa, los profesores y las condiciones de aplicación del examen no variaron, el nivel de lengua inicial de estos alumnos marcó una diferencia significativa. Especialmente es importante considerar que no todos los alumnos que recibe la UAA han cursado la asignatura de inglés como lengua extranjera en los ciclos medio y medio superior. Algunos alumnos tienen el primer contacto con el inglés precisamente al ingresar a la Universidad. Podríamos concluir que 400 hrs. de clase no fueron suficientes para alcanzar de manera

---

<sup>15</sup> Wood, *Op Cit.*, p. p 117-130.

satisfactoria el nivel *Intermediate-Mid* para esta población de alumnos, lo que nos lleva a reconsiderar algunas propuestas del Programa Lower.

Esta relación del tiempo del que se dispone con los objetivos de aprendizaje, nos lleva a apoyar la hipótesis anterior en cuanto a la insuficiencia de horas se refiere. Sumando los puntos anteriores y los resultados del examen Lower podemos hacer las siguientes recomendaciones :

- Es necesario hacer una reorganización de los objetivos de aprendizaje en cada curso.
- El Programa de comprensión de lectura está suficientemente detallado para el Nivel 1, pero sería conveniente ampliarlo para los otros tres niveles. Igualmente será necesario reorganizarlos incluso desde el Nivel 1.
- El Programa de escritura necesita ser revisado por las razones expuestas en las conclusiones del examen. Probablemente necesitará orientarse hacia otros objetivos y por lo tanto a tareas diferentes.
- Probablemente será necesario aumentar el número de horas del Programa completo, al ser evidente que se necesita mayor número de horas para cumplir los objetivos fijados. Sugerimos aumentar el número de horas porque no creemos que la solución alternativa sea modificar y bajar a los objetivos. La Universidad misma se ha fijado formar profesionistas preparados para sus actividades en el mercado de trabajo, y esto incluye las lenguas extranjeras. Creemos que el objetivo terminal del Programa Lower es el nivel mínimo de inglés que un profesionista necesitará en un puerto cuya actividad principal es el turismo.

Desde mi punto de vista, la solución estará en reorganizar el Programa de manera que, por un lado se incrementen las horas de clase y por otro los objetivos de aprendizaje sean organizados de otra manera. Por ejemplo, para el Nivel 1 sería recomendable reducir los objetivos de comprensión de lectura, e incluir otros objetivos funcionales y nocionales.

Las sugerencias anteriores me llevan a sugerir una *reconsideración* a los materiales de enseñanza utilizados hasta ahora. En este sentido será necesaria una evaluación cuidadosa del libro de texto que se ha utilizado. Propongo además la creación de un cuaderno de trabajo para los niveles 3 y 4 orientado hacia la práctica de estrategias de examen. Este material podría ser utilizado de manera individual o como complemento para los niveles mencionados. Este tipo de apoyo didáctico sería de gran utilidad para alumnos y profesores.

Finalmente quisiera concluir con la siguiente reflexión : en la medida en que se continúe el trabajo de investigación sobre el examen, aplicando éste y otros modelos de análisis, estos beneficios redundarán en todo el proceso educativo, tanto en beneficio de los mismos estudiantes con respecto al examen, como en la planeación educativa del CLE, cuyo objetivo y razón de ser es la preparación de los estudiantes en lengua extranjera para enfrentar la demanda que impone ser un profesionista en el complicado México de hoy.

Durante mi trabajo en la UAA he tenido la oportunidad de vivir y experimentar el proceso de planear, diseñar, implementar y evaluar los resultados de una parte del “fenómeno educativo”, en palabras de Hilda Taba. Durante ese tiempo, no sólo apliqué los conocimientos adquiridos con anterioridad, sino que ha significado una continua búsqueda de soluciones y alternativas, un continuo aprendizaje y maduración no sólo como profesionista, sino como persona cuyo objeto de estudio y trabajo son otros seres humanos.

De ahí ha surgido mi preocupación por estos aspectos que conforman nuestro trabajo, especialmente la evaluación y el diseño de instrumentos que nos permitan un acercamiento a la realidad de nuestros alumnos y a sus necesidades en cuanto a las lenguas extranjeras , que es nuestra razón de ser.

## BIBLIOGRAFÍA

- Alderson, J, Charles *et. al. Language Test Construction and Evaluation* Cambridge University Press. Great Britain , 1996.
- Baker, David. *Language Testing . A Critical Survey And Practical Guide*. Edward Arnold, London, 1989.
- Davies, Alan. "The Construction of Language Tests" en *The Edinburgh Course in Applied Linguistics Vol 4 . Tesing and Experimental Methods*. Oxford University Press, Great Britain, 1977.
- Gutiérrez Silvia "El enfoque psico-sociolingüístico de la acreditación evaluación UAM Xochimilco.
- Jacobs, Holly L. *et. al. Testing ESL Composition : A Practical Approach*. Newbury House, Rowley, 1981.
- Morán Oviedo, Porfirio. "La evaluación de los aprendizajes y sus implicaciones educativas y sociales" en *Perfiles Educativos* No. 13, julio-agosto, 1981, CISE, UNAM .
- Heaton, John B. *Writing English Language Tests*. Longman, Great Britain, 1975.
- Hughes, Arthur. *Testing for Language Teachers*. Cambridge University Press, Great Britain, 1990.
- Morrow, Keith "Communicative Language Testing : Revolution or Evolution ?" en *The Communicative Approach to Language Teaching*, Brumfit and Johnson editors. Oxford University Press, London 1979.
- Oller, John W. Jr "Discrete-point Tests vs. Tests of Integrative Skills" en *Focus on the Learner* Longman, London, 1975.
- Oller, John W. Jr. And Khan Frida. "Is There a Global Factor of Language Proficiency ?" en *Directions in Language Testing*, John Read editor. Singapore University Press, Singapore, 1980.
- Ordaz, Margarita. *Proyecto de creación de la D L . E*. Universidad Americana de Acapulco, Acapulco, 1992.
- Oskarsson, Mats. "subjective and Objective Assessment of Foreign Language Performance" en *Directions in Language Testing*. John Read editor. Singapore University Press, Singapore, 1980.

Popham, W. James. *Modern Educational Measurement : A Practioner's Perspective*. Allyn and Bacon, Boston. 1990.

Underhill, Nic. *Testing Spoken Language*. Cambridge University Press, Great Britain, 1982.

Wood, Robert. *Assessment and testing*. Cambridge University Press, Great Britain, 1991.

UAA. *Estatuto Orgánico*. Acapulco, Gro. 1992.