# Universidad Autónoma de Baja California

Instituto de Investigación y Desarrollo Educativo

Doctorado en Ciencias Educativas

"Evidencias de Validez del Examen de Egreso del Idioma Inglés (EXEDII"

TESIS

Que para obtener el grado de

***DOCTORA EN CIENCIAS EDUCATIVAS***

Presenta

Virginia Velasco Ariza

Ensenada Baja California, México. Mayo de 2008

# Universidad Autónoma de Baja California

Instituto de Investigación y Desarrollo Educativo

Doctorado en Ciencias Educativas

"Evidencias de Validez del Examen de Egreso del Idioma Inglés (EXEDII"

TESIS

Que para obtener el grado de

***DOCTORA EN CIENCIAS EDUCATIVAS***

Presenta

Virginia Velasco Ariza

APROBADO POR:

Dra. Norma Larrazolo Reyna
Directora de tesis

Dr. Eduardo Backhoff Escudero
Sinodal

Dra. Graciela Cordero Arroyo
Sinodal

Dra. Guadalupe López Bonilla
Sinodal

Dr. Guillermo Solano Flores
Sinodal

Dr. Manuel Jorge Gonzalez Montesinos
Sinodal

Ensenada B.C., México. Mayo de 2008

# AGRADECIMIENTOS

A Francisco, mi esposo, amigo, compañero y cómplice de mis sueños; mi apoyo para alcanzar las metas que me he propuesto y mi consuelo y aliento en mis fracasos.

A Ursula, Heleni y Francisco, mis queridos hijos porque han sido la mayor alegría y el más importante reto de mi vida.

A mi padre, Dr. Joel Velasco Bolado, por enseñarme con su ejemplo, a creer en la honradez, la justicia y el amor al trabajo y a la familia.

A mi madre, Matilde Ariza García Figueroa, quien hasta su último aliento pudo transmitirme su cariño incondicional, su buen humor y su entereza para enfrentar la vida.

A mis hermanos Joel, Laura, Leonora, Liliana y Mónica, porque compartieron conmigo lo más valioso que una persona puede tener: una familia sana y feliz.

A los seis miembros de mi comité de tesis, particularmente Eduardo Backhoff Escudero y Graciela Cordero Arroyo, quienes siempre me demostraron su confianza en mis capacidades, pero supieron señalar mis errores, para que pudiera aprender de ellos.

A los profesores del Programa del Doctorado: Carlos Topete, José María Garduño, Guadalupe López y Graciela Cordero, por compartir con nosotros, sus alumnos, los conocimientos y el compromiso que los ha hecho ser quienes han llegado a ser.

A los profesores Jesús Galaz y Manuel J. González Montesinos por ayudarme primero, a perder el miedo a las Estadísticas y después enseñarme a disfrutarlas. A Edna Luna Serrano, coordinadora del Programa del Doctorado, por su comprensión y su apoyo.

A mis compañeros Javier Organista, Sergio Pou y Cecilia Osuna, les agradezco la oportunidad de compartir con ellos los momentos más agradables de estos años de estudio. A Esperanza Viloria, Lilia Martínez y Maximiliano Cervantes, por su compromiso y su voluntad de lograr el sueño que compartimos juntos.

A todos mis compañeros del IIDE, quienes generosamente compartieron conmigo lo mejor de ellos. Les estaré agradecida siempre.

Particularmente a Javier Organista, a Maricela López Ornelas y a Ma. Alejandra Sánchez Vázquez, por su solidaridad y su amistad.

# CONTENIDO

# LISTA DE TABLAS

# LISTA DE FIGURAS

# 1. INTRODUCCION

El trabajo de investigación que aquí se presenta es el resultado de una serie de estudios realizados para aportar evidencias de la validez del Examen de Egreso del Idioma Inglés (EXEDII), construido en el Instituto de Investigación y Desarrollo Educativo (IIDE), dentro del marco del Proyecto de Investigación denominado Examen de Inglés de Egreso del mismo instituto. Este examen constituye una de las opciones de certificación del manejo del inglés, como lengua extranjera que ofrece la Universidad Autónoma de Baja California (UABC), a través de la Facultad de Idiomas.

Si se considera que el objetivo de los exámenes educativos es separar a los estudiantes en dos grupos: los que tienen un nivel determinado de aprendizaje y los que no pueden demostrarlo podrá comprenderse la necesidad de tomar las medidas necesarias para asegurar que el criterio con el que se efectúa esa separación es suficientemente objetivo, adecuado y toma en cuenta las principales variables que pueden afectar el puntaje que arroja. Dicho en otros términos, es indispensable asegurar la validez de la medición que se realiza mediante el examen.

## 1.1 Presentación del tema de tesis.

El problema que plantea la construcción de instrumentos de medición que evalúen el aprendizaje obliga a las instituciones educativas y a los docentes a poner atención en los factores relacionados con la certeza de la medición.

La medición psicológica desarrollada durante el siglo XX y posteriormente la evaluación educativa, con su rigurosa metodología derivada de la Psicometría ha producido estándares y recomendaciones nacionales e internacionales (APA, AERA, NCME, 1954; 1966; 1999) a los cuales deberán apegarse los autores de los instrumentos de medición psicológicos y educativos. Estos lineamientos y el constante esfuerzo de los investigadores han contribuido a mejorar la calidad de los instrumentos y la pulcritud de su aplicación, así como la fundamentación

adecuada para la interpretación de los resultados obtenidos, con lo cual se aumenta la certeza de que efectivamente los exámenes midan lo que pretenden medir, es decir tengan validez.

En las últimas dos décadas el desarrollo de la tecnología, particularmente la de los exámenes en formato computarizado ha permitido la aplicación de exámenes a grandes poblaciones de sustentantes en condiciones estándar. Por otro lado la aparición en el mercado de programas de cómputo para realizar análisis estadísticos complejos a través de la instrumentación de algoritmos ha contribuido al rigor de los análisis de los datos derivados de las aplicaciones masivas (Martínez Rizo, 2001).

Por lo anterior resulta oportuno hacer un esfuerzo por lograr mejores exámenes educativos, ya sea que el impacto de las pruebas sea mínimo en la vida de unos pocos estudiantes, o que afecte sustancialmente a sectores amplios de la población.

### 1.2. Organización del trabajo de tesis.

Con la finalidad de facilitar la lectura de esta tesis se ofrece a continuación la información necesaria para ubicar la información ofrecida en los diferentes capítulos del trabajo. En este capítulo 1 se plantea el problema de la investigación, con las preguntas que sirven de guía para la resolución de la problemática planteada. Se explicitan los supuestos así como las actividades de investigación redactadas en términos de objetivos y se explican las razones que justifican el desarrollo del estudio de validación del EXEDII.

El estudio se ubica en el contexto de la evaluación educativa, particularmente en el ámbito de la medición estandarizada, lo que significa que las condiciones de su aplicación y calificación deben ser equivalentes en todas las situaciones y que la población a la que se aplica comparte características esenciales, claramente definidas.

Los antecedentes más relevantes de la evaluación educativa del tipo estandarizada, así como la explicación de los exámenes orientados a un criterio son temas que proporcionan una

caracterización del tipo de examen que es el EXEDII, por lo que en el capítulo 2 se aporta una contextualización de la temática de la tesis.

En el capítulo 3 se presenta toda la información acerca del EXEDII, que es el instrumento evaluado. Se describe el proceso de su construcción y la metodología que se siguió en su construcción y el diseño de su estructura. Se ofrece también información acerca de las condiciones de su aplicación y los recursos que se necesitan para ello.

El Capítulo 4 presenta la fundamentación teórica del tema de la validez de los exámenes, que es el sustento de los objetivos y supuestos de este trabajo de tesis. La aproximación teórica de la investigación que se reporta en esta tesis supone que la validez de los instrumentos de medición educativa es un concepto unitario que se refiere a las interpretaciones que se hacen con base en los resultados obtenidos en su aplicación. Dado que el propósito del EXEDII es la certificación del manejo del inglés como lengua extranjera al nivel intermedio, para los estudiantes que egresan de la UABC, las evidencias de validez sustantiva del examen tienen que ver con las características de la prueba como instrumento de medición del constructo y con el contenido de la prueba, más aún si se considera que el EXEDII es un examen orientado a un criterio. Por lo tanto el acopio de las evidencias de validez del EXEDII se hizo a través de dos vías metodológicas distintas, reportadas de manera separada.

Por tratarse de actividades seriadas con propósitos, metodología y análisis de resultados orientados hacia distintos tipos de evidencias en el capítulo 5 se describe el proceso de búsqueda de evidencias de validez de contenido y el capítulo 6 se dedica a la indagación de las evidencias de validez de constructo.

En el capítulo 7 se hace un análisis de los resultados de las dos vías de búsqueda de evidencias de validez, intentando una interpretación integral, que trascienda los hallazgos separados, a la luz de la teoría sustantiva. El eje que articula los resultados de la investigación es la utilidad de la prueba, de acuerdo con los criterios propuestos por Bachman y Palmer (1996).

## 1.3 Planteamiento del problema.

Utilizar los puntajes obtenidos en exámenes de idiomas como base para hacer interpretaciones acerca de las habilidades que los sustentantes tienen en la lengua evaluada implica demostrar hasta qué punto la ejecución en la prueba se relaciona con el uso del lenguaje, en situaciones naturales.

De acuerdo con la literatura sobre el estudio del lenguaje éste ha sido abordado desde diferentes perspectivas que varían entre dos posturas antagónicas: las que consideran que el lenguaje es una capacidad específica, hereditaria, determinante del desarrollo del lenguaje simbólico y las que suponen que es el ambiente el que moldea la habilidad a través de la convivencia con una comunidad hablante.

Como se verá en el apartado 3.3, la metodología orientada a un criterio permite elaborar exámenes con mayores garantías de validez desde el proceso de su construcción, debido a que es primordial la definición de los criterios contra los que se compara la ejecución del sustentante. La definición operacional de las conductas que el sustentante deberá emitir para demostrar que cumple con el criterio debe derivarse de la selección cuidadosa de los objetivos que representan el dominio a medir y una vez seleccionados, se requiere definir clara, precisa y específicamente la muestra de ese dominio que deberá medir cada uno de los reactivos que se diseñen. Pero la definición de un dominio que implica el uso del lenguaje hace de la definición mencionada arriba, una tarea especialmente compleja. Baste considerar que la lengua es una característica definitoria del ser humano, que está presente en todos los procesos cognoscitivos, emocionales y sociales de los individuos y que, puesto que se adquiere a través del contacto con una comunidad parlante, ésta transmite junto con la lengua sus creencias, valores y cultura en general. Por lo anterior, en el proceso de validación del EXEDII fue necesario recurrir a un modelo de construcción y evaluación de pruebas de lenguaje que, desde un enfoque comunicativo permite organizar tanto la búsqueda de evidencias de validez del EXEDII, como la interpretación de los datos obtenidos.

El modelo que se adoptó es el de Bachman y Palmer (1996), el cual se aborda con detalle en el apartado 5.3 de esta tesis. La adopción de este enfoque permitió plantear una metodología que complementa los procedimientos que tradicionalmente se utilizan para ofrecer evidencias de validez de contenido y de constructo, integrándolos en un marco conceptual en el que se pueden tomar en cuenta todos los elementos relevantes de la medición del manejo de la lengua. Los elementos a evaluar pueden ser definidos operacionalmente y valorados en cada caso, dependiendo de la relevancia que adquieren en diferentes momentos de la vida de una prueba.

Bachman y Palmer dicen que desde la planeación, el diseño, la construcción y el uso de una prueba, hasta la interpretación de sus resultados, la cualidad más importante de un instrumento que mide el uso del lenguaje es su utilidad y proponen seis indicadores para valorarla. Enseguida explicaremos cómo se aplican esos indicadores en este estudio.

La confiabilidad puede ser evaluada en dos vertientes: desde el punto de vista psicométrico, aplicando pruebas estadísticas necesarias para demostrar la estabilidad de la prueba y de sus reactivos. También se evalúa la confiabilidad de cada reactivo a través del juicio de expertos que determinan el grado en el que éstos constituyen elementos que conforman una muestra del dominio a medir. Este juicio también aporta evidencia de validez de constructo.

La validez de constructo puede abordarse desde dos enfoques el estadístico, buscando la estructura interna de la prueba y mediante el juicio de jueces expertos, que califiquen si los reactivos miden el uso del lenguaje para el cual fueron diseñados.

La autenticidad y la interactividad se evalúan de acuerdo con la opinión de expertos quienes juzgan las tareas de la prueba de acuerdo con su similitud con a las tareas del uso del lenguaje en condiciones reales, en tres aspectos: temática, contexto cultural y lenguaje.

El impacto de la prueba puede evaluarse en dos campos: el micro, que se refiere a las consecuencias que tiene el resultado del examen para el sustentante y el macro que se refiere a las consecuencias de interpretar los resultados del examen en el ámbito de los asuntos

relacionados con las políticas educativas, asignación de recursos económicos o comparación de ejecuciones entre escuelas, estados o países. En el caso de los estudios de validación del EXEDII el nivel micro es el que resulta pertinente para valorar la utilidad del EXEDII.

Los aspectos prácticos de la prueba pueden ser abordados desde la perspectiva del sesgo que podrían producir algunas características de la prueba, o de las condiciones de su aplicación que representaran la ocasión para la respuesta diferencial de grupos de sustentantes, otorgando ventaja a unos en detrimento de otros. La opinión de los jueces puede arrojar luz sobre aspectos de este tipo.

### 1.4 Preguntas de investigación.

Con base en lo anteriormente dicho, se plantearon las siguientes preguntas para guiar la investigación que aquí se reporta:

¿Cuáles son las características del dominio de uso de la lengua meta (DULM) que debería medir el EXEDII?

¿Cuáles son las características de las tareas del dominio de uso de la lengua meta (DULM) que debería medir el EXEDII?

¿Los reactivos del EXEDII constituyen una muestra del uso del inglés de los estudiantes que egresan de la UABC y si ese es el caso, de qué manera estos datos aportan evidencias de validez de contenido y de constructo de la prueba?

¿Qué evidencias se tienen de la Utilidad de la prueba?

¿En qué grado el EXEDII constituye una escala de reactivos que miden conjuntamente un constructo, o dimensión y de qué manera estos datos aportan evidencias de validez de constructo?

¿En qué grado la estructura factorial del EXEDII es congruente con el diseño de la prueba y de qué manera estos datos aportan evidencias de validez de constructo?

## 1.5 Supuestos.

Congruente con lo anterior la presente investigación parte de los siguientes supuestos:

- Los reactivos del EXEDII evalúan los conocimientos, habilidades, o competencias representativas del nivel intermedio del uso del inglés, en los estudiantes que egresan de la UABC.
- Las tareas del EXEDII representan muestras del dominio de uso de la lengua de los estudiantes que egresan de la UABC, y las características de las mismas son congruentes con las de ese dominio y con las de los sustentantes del EXEDII.
- Los reactivos del EXEDII miden en su conjunto un constructo y las áreas en las que se agrupan según su diseño (comprensión auditiva, gramática y lectura) pueden considerarse subescalas del instrumento que miden habilidades específicas dentro del dominio evaluado.
- El EXEDII es un instrumento de medición útil para certificar el uso del inglés como lengua extranjera de los estudiantes que egresan de la UABC.

## 1.6. Objetivos.

Para investigar el grado de certeza de los anteriores supuestos se planearon los siguientes objetivos:

**Objetivo general.** Reunir evidencias de la validez del EXEDII mediante la búsqueda concatenada de elementos que aporten información sobre su validez de contenido y de constructo.

**Objetivos específicos.**

1. Recabar las opiniones de expertos en docencia al nivel superior, y en docencia del inglés como lengua extranjera sobre los conocimientos, habilidades y competencias que deberían demostrar los estudiantes que egresan de la UABC.
2. Construir una conceptualización del dominio de uso de la lengua meta (DULM) de los estudiantes que egresan de la UABC, para que pueda ser utilizada como criterio para comparar las características y contenido de los reactivos del EXEDII.
3. Comparar los reactivos del EXEDII con la conceptualización elaborada para determinar: a) si a juicio de los expertos el contenido del examen representa un dominio definido como Dominio de Uso de la Lengua Meta (DULM).
4. Analizar los datos obtenidos de las aplicaciones del EXEDII con técnicas estadísticas multivariadas para: a) determinar si todos sus reactivos miden una sola dimensión y b) indagar si las relaciones entre sus variables pueden ser explicadas por un número reducido de variables latentes, congruentes con el diseño del instrumento.
5. Interpretar los resultados obtenidos para determinar las evidencias de la utilidad de la prueba

## 1.7. Justificación.

La UABC como institución de educación superior recomienda que la formación profesional de los alumnos que egresen de las carreras ofertadas por ella incluya el manejo de una lengua extranjera. Para ese fin, el currículo de cada carrera puede incluir el aprendizaje de un idioma extranjero, o bien los alumnos pueden optar por tomar los cursos que ofrece la Facultad de Idiomas de la propia universidad. Por otra parte, para acreditar el manejo del inglés como lengua extranjera la universidad ofrece también varias opciones, una de las cuales es presentar el EXEDII.

El EXEDII es un examen estandarizado, porque es una prueba que se construyó, se aplica y se califica en condiciones uniformes. Los datos de su aplicación a lo largo de ocho años

se han analizado y se han efectuado estudios para responder a las recomendaciones de los estándares nacionales e internacionales de calidad de las pruebas estandarizadas. Siendo la validez el criterio más importante de la calidad de cualquier instrumento de medición, y aunque la validez de los exámenes criteriales se construye al momento de su construcción (Popham, 1994) los expertos en evaluación educativa en general, así como de evaluación criterial en particular recomiendan reunir todas las evidencias de validez que sea posible obtener, para garantizar la pertinencia de las interpretaciones y fundamentar las decisiones que se toman a partir de los resultados de los exámenes.

Por las razones anteriores y porque la conducción de estudios de validez es una de las tareas recomendadas por los estándares nacionales e internacionales para la construcción y uso de pruebas psicológicas y educativas, se realizó esta investigación cuyos resultados se reportan en esta tesis. Se espera que los hallazgos de estos estudios beneficien a la UABC, a los estudiantes que egresan de ella y a la comunidad en la que se insertarán los futuros profesionales, pues aportará evidencia sobre la calidad de un instrumento de evaluación de uso continuo por los estudiantes de la propia universidad.

# 2. ANTECEDENTES Y CONTEXTO

La evaluación educativa constituye el último paso del proceso de la enseñanza escolar formal. El aprendizaje de los alumnos es un proceso que se intersecta con el anterior, generando un universo de aspectos evaluables, que se abordan desde diversas perspectivas y con muy distintos procedimientos a fin de realimentar todas las etapas de ambos procesos. Los resultados obtenidos en los exámenes escolares permiten realizar interpretaciones sobre los sustentantes, así como también sobre todas las variables del ámbito de la escuela. Los maestros, los métodos de enseñanza, las características de la escuela, las condiciones socioeconómicas y psicológicas de las personas involucradas son todos factores que de una u otra forma influyen en las interpretaciones que pueden hacerse de los resultados. Por ello es necesario definir con claridad el propósito de cada examen y asegurarse de que las tareas del mismo representan el dominio de conocimientos o habilidades que se desean evaluar.

De acuerdo con Martínez Rizo (2004) en la escuela moderna la currícula consiste en una trayectoria graduada, con un principio y un término definidos en los que se requiere la presentación de evidencia de la adquisición de competencias específicas para acceder al siguiente grado. La sistematización de la obtención de la evidencia dio lugar al desarrollo de los exámenes, que hoy día constituyen la forma más aceptada para verificar el logro de competencias, el aprendizaje de conocimientos, o el desarrollo de habilidades meta primordial de cada grado escolar.

Hoy en día, en casi cualquier escuela se aplican diversos tipos de pruebas y los resultados de éstas constituyen la base sobre la cual se toman decisiones. Las decisiones pueden estar relacionadas con un alumno, o un grupo en particular; con la escuela en su conjunto y más allá de ese ámbito con aspectos como el presupuesto que los gobiernos asignan a la educación, o con políticas educativas que afectan a grandes grupos de estudiantes. Por tanto, es importante entender la evaluación educativa en un contexto en el que se consideren aspectos que trasciendan la asignación de calificaciones a partir de la aplicación de exámenes, sin asegurar que las interpretaciones que se toman con base en los resultados obtenidos son válidas. Por ello es necesario garantizar, en la medida de lo posible que las pruebas o exámenes

educativos, entendidos éstos como instrumentos de medición, sean confiables y válidos. El estudio de validación del EXEDII tiene el propósito de atender a la mencionada consideración.

## 2.1 Estudios antecedentes

Es un hecho que la evaluación educativa moderna ha tenido un desarrollo estrechamente ligado a la Psicometría, toda vez que esta disciplina tiene como objeto de estudio la medición de las así llamadas capacidades intelectuales entre las que se encuentran el aprendizaje, el pensamiento, o el juicio. Como resultado de la investigación educativa existen muchas aproximaciones a la evaluación de la educación y no hay un acuerdo entre los expertos acerca de cuál es la mejor manera de medir lo que los alumnos aprenden en las escuelas.

Los países desarrollados han realizado importantes investigaciones sobre el tema de la evaluación educativa y los hallazgos de esas investigaciones han dado lugar a políticas educativas estatales o nacionales que, a su vez han provocado reacciones encontradas por parte de los expertos, los profesores, los padres de familia y en una palabra, los contribuyentes, que son quienes pagan la educación en la mayor parte de los países (Popham, 1990). México ha seguido la pauta marcada por esas investigaciones y sucesos, de manera que a partir de la segunda mitad del siglo XX se ha dado importancia creciente a la evaluación educativa, promoviendo la objetividad y el rigor científico de los exámenes para aumentar la certeza acerca de que los medios utilizados para evaluar el aprendizaje de los alumnos, ciertamente miden lo que pretenden medir (Martínez Rizo, 2004).

En la región norte del país, la Universidad Autónoma de Baja California (UABC) ha sido pionera en el desarrollo de los exámenes elaborados de acuerdo con estándares internacionales, particularmente en lo que toca a los exámenes estandarizados que se utilizan en la actualidad para evaluar el aprendizaje de grupos numerosos de estudiantes. Este tipo de pruebas se conoce con el nombre de exámenes estandarizados. En el siguiente apartado se ofrece información al respecto para contribuir a la comprensión de la naturaleza del EXEDII como instrumento de evaluación estandarizado.

## 2.2 Exámenes estandarizados

Los exámenes estandarizados responden a la necesidad de medir a personas con características diversas con un mismo instrumento, en una aplicación esencialmente similar y con idéntica forma de calificar para poder tomar decisiones que afectan a todos los participantes. En principio parece una manera contradictoria de hacer las cosas, pero este tipo de evaluaciones pueden ser válidas y pertinentes cuando se toman en consideración los factores que podrían producir sesgo en la medición, derivado precisamente de la diversidad de los sustentantes. Los exámenes estandarizados no son un concepto moderno, desde tiempos muy lejanos y con propósitos diferentes, ha sido necesario distinguir a aquellos individuos que son capaces de hacer algo, de los que no lo son y para lograrlo se ha recurrido a ingeniosas formas de evaluación. Popham (1990) hace un relato de algunas de estas formas, utilizadas a través de la historia. En su narración dice que desde hace al menos 22 siglos se han venido elaborando pruebas que se administran, se califican y se interpretan de una manera estándar, lo que permite tomar decisiones más justas acerca de los sustentantes, disminuyendo las consecuencias adversas de esas decisiones. Técnica y funcionalmente éstos son los motivos por los que se construyen los exámenes estandarizados.

Los exámenes estandarizados se aplican a poblaciones numerosas y a menudo se utilizan para decidir con base en los puntajes obtenidos por los alumnos que los contestan, el ingreso o egreso a instituciones educativas, la certificación de estudios, o el establecimiento de políticas educativas. Ejemplos de este tipo de pruebas son muchos de los exámenes de admisión a las universidades, a los estudios de postgrado, o los numerosos exámenes que se aplican en los Estados Unidos de Norteamérica (EUA) antes de que los alumnos puedan graduarse del bachillerato. Otro ejemplo es la Prueba de Aptitudes Escolares (*Scholastic Aptitude Test (SAT)* que es un examen estandarizado, objetivo, utilizado por la mayor parte de las universidades estadounidenses para propósitos de admisión. Esta prueba es administrada desde 1926 por el Servicio de Evaluación Educativa *(Educational Testing Service, ETS)* y ha servido de modelo para la construcción de muchos otros instrumentos de evaluación educativa estandarizada.

En los últimos años se han utilizado exámenes estandarizados para evaluar y comparar el aprovechamiento de grandes grupos de alumnos, en varios países diferentes. Los estudios se centran en dos vertientes: indicadores y evaluaciones. En el rubro de los indicadores, organismos como la Organización para la Cooperación y Desarrollo Económico (OCDE), o el llamado proyecto Indicadores de los Sistemas Nacionales de Educación (ISNE) de EUA y de otros países desarrollan medidas comparativas, relacionadas con todos los aspectos de la educación. En cuanto al aspecto de la evaluación destacan el programa Tendencias en Estudios en Matemáticas y Ciencias conocido como *TIMSS (Trends in Mathematics and Science Studies)* y el programa Progreso Internacional en Lectura y Literacidad denominado *PIRLS por sus siglas en inglés (Progress in International Reading and Literacy Study)* ambos organizados por la Asociación Internacional para la Evaluación de los Logros Educativos (Internacional Association for the Evaluation of Educational Achievement, IEA, por sus siglas en inglés). Por otra parte los programas de la OCDE conocidos como PISA (*Program for International Student Assessment)* y *ALL (Adult Literacy and Lifeskills Survey)* proporcionan información sobre la calidad de la educación en los países que participan en ellos. En el ámbito de la evaluación de la educación en los países de Latinoamérica destacan las pruebas del Laboratorio Latinoamericano de Evaluación de la Calidad Educativa en las que México ha participado.

Todos los programas mencionados y otros que no se incluyen aquí proporcionan información que se utiliza para la investigación educativa con el objetivo de fundamentar las políticas educativas como las de apoyo a los maestros, u otras de interés para el público en general. Diferentes países incluyendo a México participan en este tipo de evaluaciones internacionales y cada país participante se apega a las recomendaciones de selección de las muestras y aplicación de los instrumentos estandarizados para asegurar que los resultados sean válidos y puedan ser generalizados a poblaciones mayores. En México, el Instituto Nacional para la Evaluación de la Educación (INEE) es el organismo encargado de organizar, instrumentar e interpretar los resultados de las evaluaciones internacionales desde su creación en agosto de 2002.

Antes de la creación del INEE México tenía alguna experiencia en la construcción y uso de pruebas estandarizadas. El primer ejemplo de este tipo de pruebas lo constituye el de la

Facultad de Medicina de la Universidad Nacional Autónoma de México (UNAM) elaborado en la década de los setenta. Se trataba de una prueba de opción múltiple para seleccionar a los miles de alumnos que solicitaban entrar a esa carrera. En 1992, la Universidad Autónoma de Baja California (UABC) desarrolló un examen de conocimientos y habilidades básicas, el EXHCOBA (Backhoff y Tirado, 1993) con el propósito de seleccionar a los alumnos que deseaban ingresar a esa universidad. Pocos años después, en 1994 se creó el Centro Nacional de Evaluación para la Educación Superior (CENEVAL) que ha desarrollado entre otras, pruebas estandarizadas para seleccionar a los alumnos que ingresan o egresan de la educación media, media-superior y superior. A partir de su creación el INEE asumió la responsabilidad de ofrecer a las autoridades educativas y al sector privado herramientas idóneas para la evaluación de los sistemas educativos en lo que se refiere a educación básica (preescolar, primaria y secundaria) y media superior.

El tema de la evaluación de los alumnos a través de exámenes estandarizados ha despertado polémica. Algunos de sus detractores encuentran cuestionable la utilidad de la evaluación estandarizada porque implica la demostración por parte del estudiante de competencias mínimas, que son adquiridas como resultado de un currículo que debe ser cubierto en los cursos escolares. Pero se ha observado que en ocasiones, el currículo que se cubre en la práctica docente cotidiana incluye casi exclusivamente los contenidos que van a ser evaluados en las pruebas estandarizadas. Ésto se debe a que las puntuaciones que obtienen los estudiantes en estas pruebas pueden afectar otros aspectos de la educación, como el otorgamiento de apoyos financieros y reconocimientos de valor académico (Jaeger, 1993; Popham, 1990). Por ello Martínez Rizo (2001) hace un balance de las críticas y reconocimientos que se hacen a los exámenes estandarizados, señalando las posibilidades de mejorar la metodología de su construcción, en lugar de desecharlos.

Pero la evaluación educativa no empieza, ni termina con las calificaciones que se asignan a un alumno. El proceso de enseñanza formal requiere la representación numérica en una escala determinada, de un puntaje que represente lo que el alumno ha aprendido. Como el concepto "lo que el alumno ha aprendido" es extremadamente amplio y complejo los exámenes se diseñan con un propósito, que puede ser la comparación de los resultados de los

sustentantes contra la ejecución de un grupo de sustentantes que constituye una norma comparativa, o bien contra la descripción pormenorizada de un dominio de conocimientos, habilidades o destrezas.

### 2.3 Evaluación normativa y criterial

La publicación del libro Taxonomía de los objetivos educativos de Benjamin Bloom en 1954 (Bloom, 1981) influyó de manera determinante, tanto en el desarrollo de la currícula como en la construcción de las medidas evaluativas (Hopkins, 1998). Por otra parte, la influencia de la corriente conductista en la Psicología contribuyó a la exigencia de expresar en términos conductuales y estrictamente observables la redacción de los objetivos, metas y propósitos de los programas educativos. Así, los exámenes objetivos derivados de esta aproximación se elaboraron poniendo especial cuidado en la redacción de las preguntas y de las opciones de respuesta, como es el caso de los exámenes de opción múltiple los cuales ganaron popularidad a partir de esta corriente objetivista.

De acuerdo con Nitko (1994) los puntajes obtenidos a partir de los exámenes objetivos pueden ser interpretados al menos de dos maneras: 1) se comparan con la ejecución de otros sustentantes que han presentado la misma prueba en otro momento, pero bajo circunstancias similares, cuya ejecución constituye la norma de comparación (pruebas normativas), o bien 2) se cotejan contra la descripción detallada de un dominio de ejecuciones que el estudiante debió haber aprendido como resultado de la instrucción en un currículo determinado (pruebas criteriales). La evaluación normativa proporciona información sobre el rendimiento comparativo de grupos de estudiantes aunque no precisa cuánto y qué aprendió cada individuo; y la evaluación criterial por su parte, no necesariamente sitúa la ejecución de un estudiante en relación con sus iguales, pero sí informa cuánto puede hacer y cuánto no puede hacer del contenido de un programa determinado.

Tabla 2.1. Diferencias entre pruebas normativas y criteriales. Tomado de Hambleton (1988).

| Aspecto diferente | Pruebas normativas | Pruebas criteriales |
|---|---|---|
| a) Propósito | Compara individuo *vs* grupo (norma) | Compara individuo *vs* conjunto de competencias |
| b) Grado de especificación de contenidos | Amplias. Describen dominios generales de competencias | Detalladas. Describen dominios específicos de competencias |
| c) Desarrollo de la prueba | Énfasis: descripción general del dominio y estadísticas de los resultados (índices de dificultad y de discriminación) | Énfasis: descripción puntual del dominio a evaluar y asegurar congruencia con los reactivos del examen |
| d) Generalización de los resultados | Alcance: población con características similares | Alcance: todo el dominio evaluado |

Los exámenes normativos y los criteriales difieren en varios aspectos. Hambleton (1988) los resume en cuatro características: a) el propósito para el que se usan; b) la especificación del contenido; c) el desarrollo de la prueba y d) la generalización de las calificaciones. En la Tabla 2.1 se sintetizan estas diferencias.

Derivado de lo anterior, la decisión de evaluar de acuerdo con una norma o con un criterio depende del propósito que se persigue, así como del tipo de inferencias que se harán a partir de los puntajes obtenidos y de la clase de decisiones que se desea tomar con base en ellos. Ambos tipos de exámenes proporcionan información valiosa, siempre que estén construidos rigurosamente y se usen para los fines para los que fueron hechos.

La rigurosidad en la construcción y uso de las pruebas influye en la validez de la prueba. Por eso, en cualquier instrumento de medición la validez y la confiabilidad son los dos indicadores de calidad más importantes. En general los exámenes normativos garantizan estas dos características a través de estudios estadísticos posteriores a la aplicación piloto, así como de aplicaciones sucesivas del instrumento. En las pruebas criteriales la garantía de la adecuación de los reactivos proviene de la forma en que éstos son construidos es decir, de la puntualidad con la que se siga la metodología de la construcción de exámenes orientados a un criterio. En la metodología criterial un grupo de expertos empieza por definir el constructo que va a ser medido por la prueba. Para ello como dice Popham "construyen una sofisticada red verbal"

(1990:268) que sea capaz de describir el dominio que mide la prueba. La descripción del dominio debe de ser tan precisa que pueda definir el constructo a medir, en términos de las conductas que los sustentantes deberán emitir para responder a las preguntas de la prueba. Una vez aclarado ese punto, determinan qué temas debe incluir el examen y en qué proporción; es decir, seleccionan el contenido del examen. Posteriormente otro grupo de expertos elabora las especificaciones de los reactivos de que constará el examen y un tercer grupo elabora los reactivos de acuerdo con las especificaciones. En cada uno de estos pasos se discute y se toman acuerdos de manera consensuada para asegurar que el instrumento cumpla con los requerimientos de un examen de calidad (Popham, 1990; Hambleton, 1988). Posteriormente, cuando el examen ha sido aplicado varias veces a la población real, es pertinente llevar a cabo análisis estadísticos que proporcionen información del comportamiento del examen.

Los exámenes normativos son adecuados para evaluar conocimientos y destrezas aprendidos a partir de un currículo determinado, puesto que las sucesivas generaciones de grupos de estudiantes pueden ser comparadas con la ejecución del grupo que constituye la norma. En cambio en los exámenes de certificación, la situación es diferente porque cuando se certifican conocimientos no siempre se tiene certeza acerca de cuál fue el contexto de aprendizaje de los sustentantes.

Los exámenes educativos, sean normativos o criteriales pueden ser utilizados para diferentes fines, como el diagnóstico de habilidades, la ubicación de los estudiantes en los niveles de instrucción adecuados, la selección de los mejores para el ingreso a programas o cursos, o de los menos aventajados para cursos remediales, etc. También pueden ser utilizados para certificar el manejo de un dominio determinado de conocimientos o destrezas. Los exámenes de certificación del manejo de idiomas son ampliamente utilizados en todo el mundo. Hughes (1989:9) define a los exámenes de certificación de idiomas como "pruebas que miden la habilidad de las personas en una lengua, independientemente del entrenamiento que puedan haber tenido en esa lengua". Durán (1986) afirma que las personas que presentan exámenes de certificación en una lengua no nativa pueden pertenecer a muy diversos entornos sociales, étnicos y educacionales y recomienda tomar en cuenta esta variable cuando se interpretan los resultados. Bachman y Palmer (1996) coinciden en este punto y refiriéndose al sesgo de los

instrumentos de medición en general advierten que los exámenes pueden ser válidos para una población, pero no para otra y que los exámenes de certificación de una lengua extranjera son un claro ejemplo de que los sustentantes pueden provenir de distintas geografías, culturas, razas e ideologías.

Por lo anterior parece claro que la metodología criterial es la más adecuada cuando se trata de construir un examen de certificación, ya que como se dijo arriba, la ejecución de cada sustentante no se comparará con una ejecución normativa, sino con la descripción de un dominio que se supone maneja el sustentante, por lo que no es relevante si se encuentra por arriba o por debajo de la ejecución de otros sustentantes, sino qué proporción de conocimientos o habilidades tiene de entre una muestra representativa de los conocimientos y habilidades que caracterizan a un Dominio determinado.

## 2.4 Exámenes de alto impacto

Los exámenes estandarizados pueden tener consecuencias que definen aspectos cruciales en la vida de los sustentantes como puede ser el egreso de ciertos niveles escolares, o el ingreso a instituciones de educación superior, por ejemplo. Así mismo las medidas administrativas respecto de una escuela, o de un grupo de escuelas pueden ser tomadas con base en los resultados de los exámenes estandarizados. En estos casos, se habla de exámenes de alto impacto.

La evaluación de alto impacto se refiere a una clase de exámenes que se han venido usando en varios países -- particularmente en EUA que ha sido pionero y representante de este tipo de evaluación-- cuyo propósito principal es establecer estándares de medición y de enseñanza que permitan la comparación de grandes grupos de estudiantes, como por ejemplo entre varias escuelas, o entre estados de un país, o incluso entre países. Dada la importancia de las consecuencias de estos exámenes se ha promovido el desarrollo de nuevas formas de evaluación más adecuadas y pertinentes que favorezcan decisiones más justas. Así mismo, se han aplicado algunos descubrimientos de la ciencia que han permitido el avance de la

evaluación educativa moderna. Martínez Rizo (2001) menciona los siguientes: a) técnicas para valorar el sesgo; b) reactivos que permitan no solamente la selección de la respuesta correcta (de opción múltiple) sino de respuesta construida, pero con modalidades de calificación objetiva; c) pruebas de desempeño auténtico y pruebas adaptativas por computadora; d) avances en la estadística con el análisis factorial y el muestreo matricial que permiten aplicar pruebas que cubran mejor ciertos dominios; e) en el aspecto de la administración misma de la prueba se han instrumentado procedimientos para atender las condiciones particulares de ciertos sustentantes; e) el desarrollo de pruebas orientadas a un criterio, las cuales vienen a complementar la metodología orientada a una norma.

Como los exámenes de alto impacto pueden y a menudo tienen importantes consecuencias que afectan a los individuos, a las instituciones y a la sociedad en su conjunto la Asociación Psicológica Americana *(American Psychological Association, APA)*, la Asociación Americana de Investigación Educativa (*American Educational Research Association, AERA*) y el Consejo Nacional de la Medición en Educación (*National Council on Measurement in Education, NCME)* publicaron conjuntamente (1966, 1974 y 1989) los Estándares para la Evaluación Psicológica y Educativa. El propósito de estos lineamientos es el de hacer recomendaciones para los profesionales dedicados a construir, aplicar y administrar instrumentos de evaluación para que éstos cumplan con los requisitos técnicos necesarios para favorecer interpretaciones adecuadas a partir de los resultados de los alumnos. La evaluación estandarizada en general y los exámenes estandarizados de alto impacto en particular, deben ser actividades que se emprendan con la más alta seriedad y compromiso porque como dice Jaeger (1993, P. 511) "una de las principales responsabilidades del profesional de la medición [educativa] es la de cuestionar la validez de las inferencias derivadas de los exámenes, por lo que es necesario promover el uso de prácticas de medición que incrementen la validez de esas inferencias".

El EXEDII nació como examen de alto impacto y durante diez años la aprobación del mismo constituyó un requisito reglamentario para acceder al egreso de las carreras de la UABC. Desde 2006 el requisito de acreditación de una lengua extranjera no es un requisito, sino una recomendación para los estudiantes de esa universidad. No obstante sigue siendo necesario recabar todas las evidencias que sea posible para validar las interpretaciones que se hacen de

los puntajes que los sustentantes obtienen en el EXEDII si se pretende afirmar que con ello se certifica el manejo del inglés como lengua extranjera al nivel intermedio.

# 3. EL EXAMEN DE EGRESO DEL IDIOMA INGLES (EXEDII).

En este capítulo se documentan los antecedentes y el contexto de la construcción del EXEDII con la intención de explicitar las circunstancias, los procedimientos y las características del proceso de su diseño y elaboración por considerar que esta información es importante para comprender cabalmente el objeto de estudio de la investigación que en este trabajo se reporta.

## 3.1 Desarrollo del EXEDII

Entre 1995 y 2006 la UABC exigió a sus egresados que acreditaran el manejo de alguna lengua extranjera antes de ser aceptados para su titulación. El Artículo 35 del Reglamento General de Admisión, Inscripción, Evaluación de los Alumnos y su Seguimiento en los Planes de Estudio de la UABC (1995) establecía que "Los planes de estudio de cada carrera establecerán la etapa en que los alumnos deberán acreditar el conocimiento de un idioma extranjero, por lo menos a nivel intermedio, como requisito necesario para egresar de la Universidad. La Escuela de Idiomas, en coordinación con las unidades académicas establecerá el procedimiento para la certificación de este requisito" (p. 55).

En cumplimiento del mencionado requisito la Escuela de Idiomas (EI), hoy Facultad de Idiomas de la propia universidad utilizó durante tres años y para el caso de la certificación en inglés algunos exámenes de certificación comerciales como el *Test of English as a Foreign Language (TOEFL)*.

Dado que la población de estudiantes que egresan de las carreras que imparte la UABC había venido presentando un aumento gradual considerable, la solicitud de exámenes de lengua extranjera, particularmente de inglés fue incrementándose al punto tal que la infraestructura de la EI no podía satisfacer esa demanda. La aplicación y calificación manual de exámenes generaba problemas administrativos y económicos, además de que resultaba muy oneroso para los propios estudiantes. Por esa razón, la entonces directora de la EI Mtra. Kora Basich solicitó al Dr. Eduardo Backhoff, entonces director del Instituto de Investigación y Desarrollo Educativo (IIDE) su colaboración para solucionar ese problema y tras valorar las diferentes opciones se

llegó a la conclusión de que lo mejor era construir un examen específicamente diseñado para la población de estudiantes de esta institución, que pudiera ofrecerse en formato computarizado y que cumpliera con los lineamientos internacionales de calidad técnica y con claras evidencias de calidad psicométrica, es decir, confiabilidad y discriminación adecuadas y con un nivel de dificultad intermedio, tal como lo recomendaba el mencionado reglamento.

Para poder estar en condiciones de construir un examen de certificación se abrió un proyecto de investigación en el IIDE. El Proyecto del Examen de Inglés de Egreso obtuvo financiamiento de la UABC y se invitó a investigadores a trabajar en la creación del instrumento de medición.

El plan de trabajo acordado por los investigadores del proyecto se detalla en la Tabla 3.1 y está fundamentado en las recomendaciones de Hambleton (1988), Nitko (1994) y Popham (1990). Se hicieron algunos ajustes en la secuencia de los pasos sugeridos por Hambleton y se agruparon algunos de ellos, con el fin de adaptarlos a las características particulares de este proyecto. Tal y como lo recomienda la metodología se conformó un primer comité denominado Comité Coordinador (CC) integrado por investigadores y académicos del IIDE. La responsabilidad de este comité fue la coordinación de la construcción de examen, así como la asesoría en los aspectos técnicos. Este comité se encargó de las consideraciones preliminares, que es el primer paso del plan de trabajo.

El EXEDII comenzó como un examen en formato de lápiz y papel y posteriormente se modificó su formato para aplicarlo y calificarlo en computadora. También en esta etapa se decidió que fuera de opción múltiple y orientado a un criterio ya que se trataba de un examen de certificación.

Para proceder con el punto número 2 se contrató a algunos profesores de inglés de la EI de la UABC para colaborar en distintas etapas del proyecto. Se conformó un segundo comité, denominado Comité de expertos (CE) cuyo objetivo fue revisar los contenidos del currículo que serviría como criterio. Se consultaron los programas vigentes en la Escuela de Idiomas en Mexicali, Tijuana y Ensenada y se concluyó que el dominio que debería medir el EXEDII debería

implicar las habilidades suficientes para permitir la comunicación oral y escrita que típicamente demanda el medio laboral para los alumnos egresados de la UABC. Así mismo, los expertos determinaron que el programa que en ese momento estaba vigente en Ensenada cumplía con los requisitos que la metodología recomendaba, ya que las habilidades que pretendían medir eran las que se desarrollaban en el nivel 2 del curso del programa de *ScottForesman* (1991).

Recuadro 3.1 Plan de trabajo diseñado para la construcción del EXEDII, según los lineamientos sugeridos por Hambleton (1988), Popham (1990) y Nitko (1994).

| Etapas de la construcción del EXEDII |
| --- |
| Consideraciones preliminares.<br>1.1 Especificar el propósito de la prueba.<br>1.2 Especificar los grupos a evaluar y cualquier requisito especial de la prueba.<br>1.3 Determinar el tiempo disponible para producir la prueba.<br>1.4 Identificar a los especialistas en contenido y en evaluación.<br>1.5 Estimar el tamaño de la prueba.<br>Revisión de contenidos.<br>2.1 Analizar el contenido y los objetivos a evaluar<br>2.2 Especificación de preguntas:<br>2.2.1 Descripción general breve y concisa del contenido y/o conductas<br>2.2.2 Ejemplo de las indicaciones de la prueba y un modelo de pregunta<br>2.2.3 Atributos de la pregunta. Especificación de las áreas que se incluyen y las que no se incluyen en la descripción del contenido.<br>2.2.4 Atributos de las respuestas incorrectas que se deberán elaborar. La estructura y contenido de éstas, deberá ser tan detallada como sea posible.<br>3. Elaboración de preguntas.<br>3.1 Elaborar un número suficiente de preguntas para la prueba piloto.<br>3.2 Editar las preguntas elaboradas.<br>4. Evaluación de la validez de contenido.<br>4.1Revisar las preguntas de la prueba para determinar si cumplen con las especificaciones de contenido.<br>Revisar las preguntas para determinar su adecuación técnica.<br>Con base en los dos puntos anteriores, corregir las preguntas o cambiarlas.<br>Escribir preguntas adicionales (si es necesario) y repetir el paso 4.<br>5. Administración de la prueba de campo.<br>Organizar las preguntas de la prueba en un formato para la prueba piloto.<br>5.2 Administrar el formato de prueba a grupos seleccionados adecuadamente.<br>5.3 Realizar el análisis de las preguntas, estudios de validez y estudios de sesgo (tendencias inadecuadas de las preguntas).<br>Corregir o eliminar las preguntas de la prueba cuando se juzgue pertinente, con base en los resultados del punto anterior.<br>6. Ensamble de la prueba.<br>6.1Determinar el tamaño de la prueba, el número de formatos que se necesitan y el número de preguntas por objetivo.<br>Seleccionar las preguntas de la prueba del banco de preguntas válidas.<br>Preparar las instrucciones, preguntas de práctica, formato, clave de respuestas.<br>7. Selección de un estándar.<br>7.1 Iniciar un proceso para determinar el punto de corte.<br>8. Preparación de manuales.<br>8.1 Preparar un manual para la administración de la prueba.<br>8.2 Preparar un manual técnico.<br>9. Colección de datos técnicos.<br>9.1 Conducir investigaciones de confiabilidad y validez. |

De esta manera el mencionado programa se convirtió en el criterio al que se orienta el EXEDII. Se trata de un curso de inglés diseñado para la enseñanza del inglés como lengua extranjera para alumnos adultos. Hace hincapié en el desarrollo de las habilidades comunicativas de comprensión auditiva, lectura, escritura y conversación apoyado por la adquisición de nociones de gramática. El programa consta de tres niveles de dificultad creciente, cada uno con dos subniveles a saber: *In Contact* niveles 1 y 2 (principiante); *On Target* niveles 1 y 2 (intermedio); *In Charge* niveles 1 y 2 (avanzado).

La estrategia que se utilizó para identificar los contenidos relevantes fue la que propone Contreras (1998), misma que está basada en el modelo de Nitko (1994). Brevemente, el método incluye a): la elaboración de una retícula que es un modelo gráfico que representa los contenidos de un curso, así como las relaciones de servicio que dan y reciben esos contenidos entre sí. Con base en el análisis visual de esas relaciones se puede decidir qué contenidos son esenciales (llamados contenidos esenciales) para el curso y por lo tanto deben ser evaluados en el examen. Así mismo es posible ver cuáles contenidos pueden, pero no necesariamente tienen que ser incluidos en el examen (llamados importantes) de acuerdo con una selección al azar; b) la redacción de un documento denominado Tabla de Justificación de Contenidos da cuenta de las decisiones tomadas acerca de la inclusión de los contenidos esenciales y de los contenidos importantes a los que se hizo referencia arriba; c) el siguiente paso es la elaboración de una Tabla de Especificaciones del examen, la cual resume las decisiones adoptadas en los dos documentos anteriores respecto de la planeación de la prueba. En la tabla de especificaciones se especifican los temas y subtemas del dominio que se van a evaluar, el número y las características de cada uno de los reactivos de la prueba y las características de las opciones de respuesta.

Con los contenidos y sus relaciones claramente explicitados en la retícula El CE procedió a construir la Tabla de Especificaciones del EXEDII, e hicieron una descripción detallada de cómo debería ser cada uno de los reactivos de la prueba en todas las versiones que se hagan de ella. Un requisito de la metodología es que todas las decisiones que se tomen con respecto a

la prueba sean consensuadas por lo que los contenidos incluidos en el EXEDII, así como las características de los reactivos fueron discutidas y aprobadas por consenso en los dos comités.

La metodología recomienda que el comité que elabora las especificaciones de la prueba sea diferente del que elabora los reactivos, pero dado que no había suficientes profesores de inglés que pudieran colaborar con la construcción del EXEDII se decidió que el CE fuera renovado parcialmente. Así, uno o dos de los integrantes del comité en cada sede (Tijuana, Mexicali y Ensenada) permanecieron para aprovechar su experiencia y ayudar a los nuevos profesores que se sumaron al comité en cada sede. El CE renovado se dio a la tarea de construir los reactivos del EXEDII con base en las especificaciones descritas. Se acordó que los profesores de inglés radicados en Mexicali elaboraran los reactivos de la subescala de Gramática, los profesores de Tijuana elaboraran los de Comprensión auditiva y los de Ensenada los de Lectura. Una vez concluida la primera versión completa de 100 reactivos se procedió a realizar el jueceo de los reactivos. Para evitar que los elaboradores de reactivos evaluaran sus propios reactivos se acordó que los profesores que elaboraron los reactivos de una subescala revisaran los reactivos de las otras dos. El comité coordinador (CC) revisó el trabajo de todos. Los criterios que guiaron esta revisión recomendaban cuidar que los reactivos estuvieran apegados a las especificaciones, que no hubiera evidencia de sesgo, claves inadvertidas o problemas de redacción y ortografía.

La primera versión completa del EXEDII se sometió a una prueba empírica en los tres *campi* con una muestra intencional de 30 estudiantes voluntarios del nivel intermedio (nivel 2 del curso *On Target*) por entidad, que en ese momento eran los alumnos de tercer semestre del programa de la EI. También se aplicó el EXEDII a una muestra de 30 alumnos de cada una de las entidades de Mexicali, Ensenada y Tijuana de los niveles inmediatamente inferior y superior (niveles 1 y 3 del curso), es decir de segundo y cuarto semestres. Adicionalmente se solicitó a los alumnos de quinto semestre que también contestaran el EXEDII ya que se deseaba calibrar el examen en cuanto a su nivel de dificultad, con respecto a los niveles del programa de inglés de la EI. Los datos obtenidos fueron analizados para obtener su índice de dificultad (proporción de aciertos, o valor p), su coeficiente de discriminación ($r^{bis}$) y su coeficiente de confiabilidad (alfa de Cronbach). Estos estadígrafos se usaron como criterios para editar los reactivos que

mostraran problemas técnicos. Con los reactivos editados se volvió a aplicar el EXEDII, pero esta vez a la población a la que está dirigida el examen. Se tenía previsto que los estudiantes que contestaran las primeras aplicaciones del EXEDII serían probablemente los que más seguros se sintieran de aprobarlo y por lo tanto se esperaban puntajes altos al principio, mismos que irían disminuyendo paulatinamente hasta llegar al nivel real de la población meta.

El EXEDII empezó a aplicarse periódicamente en 1999. Al principio se aplicó dos veces por año en Mexicali, Tijuana y Ensenada, pero pronto fue evidente que era necesario aplicarlo con mayor frecuencia, razón por la que se empezó a aplicar dos veces por semestre.

A lo largo de nueve años de aplicaciones del EXEDII se ha ido recabando información acerca de diversos aspectos del examen en la población real y se han realizado algunos de los estudios contemplados en el plan de su desarrollo, especificados en el Recuadro 3.1. Por ejemplo se han realizado algunos análisis sobre su calidad técnica.

En cuanto a la preparación de los manuales del EXEDII, es una actividad que está en proceso y que se planea tener completa para 2008. Finalmente los estudios de validez del EXEDII son reportados en este trabajo de tesis.

### 3.2 Recursos materiales para la administración del EXEDII.

Para el formato computarizado del EXEDII se utilizó el mismo software especializado (SICODEX) que se utiliza para el Examen de Habilidades y Conocimientos Básicos (EXHCOBA), que es el examen que la UABC utiliza para la selección de los alumnos que aspiran a ingresar a esta universidad. La implementación del EXEDII en formato computarizado se desarrolló según el ciclo de vida de desarrollo denominado "Entrega evolutiva" que significa que se crea una versión completamente operativa, a la que se le agregan los elementos que se vayan juzgando necesarios. Con este tipo de modelo se cuenta con una versión que se va refinando gradualmente, pero siempre se tienen una versión lista para entregar.

Fue necesaria la adquisición de equipo y material de computación adecuado para hacer uso del examen computarizado. Las máquinas de los centros de evaluación de la UABC, en el momento de las primeras aplicaciones del EXEDII a la población real tenían las siguientes características: sx386 con 4 mb de ram, sin disco duro, monitores a color vga, ratón, en red Novell 3.12. Corrían windows 3.1, en modo seguro (win/s). Actualmente, se han modernizado y complementado los equipos en las salas de evaluación de la Facultad de Idiomas (FI) en Mexicali y Tijuana, así como en la sala de evaluación de Ensenada que son los lugares en donde se aplica el EXEDII. Con respecto a las características físicas de los salones de evaluación se puede afirmar que son adecuadas ya que tienen un número suficiente de computadoras, cuentan con condiciones de temperatura e iluminación constantes y adecuadas.

## 3.3 Procedimiento estándar para la administración del EXEDII

Actualmente el EXEDII es aplicado dos veces al mes en Tijuana y Mexicali y de dos a cuatro veces por semestre en Ensenada. Esta frecuencia puede variar, dependiendo de la demanda. Desde el punto de vista administrativo los estudiantes que desean presentar el EXEDII solicitan su ficha y pagan el importe del examen en las oficinas de la FI en cada entidad. Se capturan en un listado el nombre, matrícula y carrera de cada estudiante y se hace llegar a los responsables de la aplicación del EXEDII en el centro de evaluación correspondiente. Una vez en la sala de evaluación se asigna una computadora y una diadema con audífonos para cada estudiante. Se les pide que capturen su matrícula y automáticamente aparecen los datos del estudiante. El encargado de la aplicación del examen solicita a los estudiantes que lean las instrucciones (en español) y que levanten la mano si tienen alguna duda. Enseguida se les da oportunidad de que escuchen el ejemplo (reactivo 101), mismo que se aprovecha para que ajusten el volumen de los audífonos. Una vez que inician el examen los sustentantes tienen un tiempo máximo de 90 minutos para resolverlo al término del cual, el examen se cierra automáticamente mientras aparece una leyenda en la pantalla en la que se les informa que se terminó el tiempo disponible para resolverlo. La mayor parte de los estudiantes terminan antes de que se cierre el examen.

Se han hecho esfuerzos por estandarizar las condiciones de aplicación del EXEDII, reduciendo al mínimo la intervención de los aplicadores. Sin embargo, la comunicación con las personas que aplican el EXEDII en cada entidad ha proporcionado información acerca de algunos aspectos del examen que resultan problemáticos, como las instrucciones y el hecho de que se escucha solamente una vez el diálogo entre dos personas, lo que constituye el estímulo para la subescala de Comprensión auditiva. Esta información sugiere que ciertas características del EXEDII deben mejorarse.

En relación con la mecánica para responder el EXEDII se puede afirmar que el formato computarizado no supone una dificultad adicional para la mayoría de los sustentantes, ya que su interfaz es amigable, aunque en una de las actividades que se organizaron para la validación de contenido algunos panelistas sugirieron hacer modificaciones a la interfaz porque resulta anticuada. Cuando el sustentante entra al examen la información necesaria se presenta en la pantalla, dividida en dos partes. En el lado derecho de la pantalla se muestra un tablero con celdillas numeradas del 1 al 101, que corresponden a las preguntas del EXEDII, más el ejemplo. El sustentante solamente tiene que señalar con el cursor el número de reactivo que desea contestar y éste aparece en la parte izquierda de la pantalla. Allí se muestra el estímulo visual que en el caso de la sección de Gramática corresponde a una frase incompleta y en el caso de la sección de Lectura consiste en un texto y una pregunta. Las opciones de respuesta de todo el examen son cuatro más la opción *I don't know*. A diferencia de las dos subescalas anteriores, en la de Comprensión auditiva solamente se proporciona un apoyo visual como parte del estímulo, el cual consiste en que el alumno puede ver la pregunta y las opciones de respuesta antes de contestar, pero el diálogo que constituye el estímulo auditivo no aparece escrito en la pantalla y además, solamente puede ser escuchado una vez. El sustentante puede contestar a las preguntas en el orden que desee, pero generalmente lo hacen de la 1 a la 100.

### 3.4 Estructura del EXEDII

La estructura del EXEDII se muestra en las Tablas 3.1, 3.2 y 3.3 en la que se detallan las tres subescalas, los nodos y objetivos. El examen consta de 100 reactivos, de los cuales 32 corresponden a Comprensión Auditiva, 34 a Gramática y 34 a Lectura.

Como puede observarse en la Tabla 3.1 la redacción de los objetivos está expresada en términos de las habilidades particulares que supuestamente mide cada reactivo. La dificultad que deberían tener los reactivos está gruesamente representada en su posición en la tabla. Es decir, los reactivos más fáciles corresponden a las tareas más sencillas, en las que el sustentante debe identificar o reconocer elementos evidentes en el estímulo. Estos reactivos en términos generales están localizados en la mitad superior de la tabla. Los reactivos que implican inferencias o razonamientos más complejos aparecen en la mitad inferior de la tabla. No obstante, dado que los nodos implican habilidades que se ejecutan con palabras, frases o párrafos pueden encontrarse reactivos con dificultad mayor en la mitad superior y viceversa.

Tabla 3.1 Estructura temática de la subescala de Comprensión auditiva del EXEDII.

| Subescala | área | Nodo | | Objetivo | | Reactivo |
|---|---|---|---|---|---|---|
| Comprensión auditiva | Identificar y comprender | Palabras | Inform.específica | Implícita | Objeto | 1 |
| | | | Signif. contextual | | | 6 |
| | | | Implicación | | | 23 |
| | | | Signif. contextual | | | 9 |
| | | Frases | Información | | | 7 |
| | | | | Específica | | 16 |
| | | | | | | 22 |
| | | | Clasificación | Noticia | Jerarquía | 13 |
| | | | | Programa | | 21 |
| | | Diálogos | Inf. Específica | Conclusión | | 10 |
| | | | | Actividad | | 17 |
| | | | | Paráfrasis | | 24 |
| | | | | Justificación | | 27 |
| | | | | Sinónimo | | 31 |
| | | | Tema general | | | 2 |
| | | | Inferir | Comportamiento | | 4 |
| | | | | Resultado | | 20 |
| | | | Hecho u | Implícita | | 11 |
| | | | Opinión | Explícita | | 19 |
| | Razonar | Frases | Info.equivalente | | | 5 |
| | | | Comparar/contrastar | | | 8 |
| | | | Inferir | Decisión | | 12 |
| | | | | Fecha | | 14 |
| | | | | Actitud | | 28 |
| | | Diálogos | Inferir | Cualidades | | 3 |
| | | | | Categ. Mayor | | 18 |
| | | | | Conclusión | | 29 |
| | | | | Consecuencia | | 32 |
| | | | | Síntesis global | | 15 |
| | | | Resumir | Idea principal | | 25 |
| | | | | Aseverar sin | | 26 |
| | | | Hecho/opinión | apoyo | | 30 |
| | | | | Hecho | | |

Tabla 3.2 Estructura temática de la subescala de Gramática del EXEDII.

| Subescala | Subárea | Objetivo | Reactivo |
|---|---|---|---|
| Gramática | Conjugaciones: concordancia de tiempo, persona, género y número | Presente: verbo, sujeto, 3a persona, singular. | 33 |
| | | Presente: verbo, sujeto, 3a persona, plural | 34 |
| | | Presente: Interrog, (be) sujeto, 3ª persona, singular | 35 |
| | | Presente: Interrog, (be) sujeto, 3ª persona, plural | 36 |
| | | Presente: Interrog, (do) sujeto, 3ª persona singular | 37 |
| | | Presente: Interrog, (do) sujeto, 3ª persona, plural | 38 |
| | | Presente: progresivo, sujeto | 39 |
| | | Presente: progresivo, interrogativo | 40 |
| | | Presente: progresivo, contraste con presente | 41 |
| | | Pasado: verbos irregulares | 47-50 |
| | | Pasado: verbos regulares | 51 |
| | | Pasado: forma interrogativa | 52 |
| | | Presente Perfecto: verbo, sujeto | 53 |
| | | Presente perfecto: Interrog : verbos regulares | 54 |
| | | Presente perfecto: Interrog : verbos irregulares | 55 |
| | | Presente perfecto: contraste con pasado simple | 56 |
| | Adverbios | Una palabra | 42 |
| | | Una palabra con verbos de acción | 43-44 |
| | | Frases adverbiales, repetición periódica | 45-46 |
| | Comparativo y Superlativo | Comparativo: Adjetivos cortos y largos | 57-58 |
| | | Superlativo: Adjetivos cortos y largos | 59-60 |
| | Pronombres relativos | Pronombres relativos: objetos inanimados | 61 |
| | | Pronombres relativos: objetos animados | 62 |
| | | Pronombres relativos: sujetos inanimados | 63 |
| | | Pronombres relativos: sujetos animados | 64 |
| | Cláusulas | Contraste de tiempos: pasado/pasado progresivo | 65-66 |

Tabla 3.3. Estructura temática de la subescala de Lectura (EXEDII).

| Area | Subárea | Nodo | Objetivo | | | Item |
|---|---|---|---|---|---|---|
| Lectura | Identificar y Comprender | Palabras | Vocabulario En contexto | Clave: | Estruct. retórica | 4 |
| | | | | Clave: | Complemento | 5 |
| | | | | Clave: | Ejemplo | 16 |
| | | | Referente | Pronombre | Personal | 21 |
| | | | | Pronombre | Relativo:*that/which* | 28 |
| | | | | Pronombre | Relativo: *that/who* | 29 |
| | | Enunciados | Compren numérica | | | 2 |
| | | | Opinión/hecho | hecho | | 9 |
| | | | Inf. específica | En: | Explicación | 8 |
| | | | | En: | Ejemplo | 14 |
| | | | | Después de | Marcador de discurso | 19 |
| | | Párrafos | Idea principal | Textual | | 1 |
| | | | Opinión/hecho | Opinión | Diferente | 15 |
| | | | Diferenciar: | Comparativo | Superlativo | 33 |
| | | | Secuencia | Marcador | De secuencia | 27 |
| | | | | Adverb/prepos | | 34 |
| | | | Inf. específica | Proposición | Adjetiva | 25 |
| | Razonar | Palabras | Vocabulario En contexto | Clave: | Comparación | 10 |
| | | | | Clave: | Palabra con afijo | 11 |
| | | | | Clave: | Sinónimo | 17 |
| | | | | Clave: | Explicación | 20 |
| | | | Marcador De discurso | Result/consec | | 30 |
| | | | | Adición | | 31 |
| | | | | Cronología | | 32 |
| | | Enunciados | Inferir | Dato | | 12 |
| | | | | Fecha | | 26 |
| | | | Contrastar inf. | Datos | Numéricos | 22 |
| | | | | Datos: | No numéricos | 23 |
| | | | Vocab en contexto | Clave | Definición/ referenc. | 18 |
| | | Párrafos | Opinión/hecho | Autor o | Terceros | 3 |
| | | | Inferencia | Actitudes | Del autor | 6 |
| | | | Idea principal | No textual | Paráfrasis 1 oración | 7 |
| | | | | No textual | Paráfrasis varias oraciones | 13 |

### 3.5 Características de los sustentantes

La población a la que va dirigido el EXEDII son los estudiantes que egresan de las carreras de la UABC, aunque a partir de 2004 se permitió que también pudieran presentarlo los alumnos de séptimo semestre en adelante. Esta medida se tomó para estar en mejor posibilidad de atender la demanda.

Los estudiantes que egresan de todas las carreras que oferta la UABC presentan características propias, derivadas de las peculiaridades de las carreras y del campo profesional en el que se ubican, pero para efectos de caracterización de los egresados en cuanto a su papel como sustentantes del EXEDII, se considera que presentan similitudes por haber estudiado en una universidad estatal, ubicada en la zona fronteriza con los Estados Unidos de Norteamérica, particularmente con el Estado de California. En su mayoría son adultos jóvenes, de nacionalidad mexicana o latinoamericana y por lo tanto su lengua nativa es el español.

# 4. MARCO TEORICO

Por muchos años la validez ha sido definida como el grado en el que una prueba mide lo que pretende medir (Anastasi, 1977). Esta es una definición que ha prevalecido quizá porque en su brevedad, se refiere al significado más esencial del concepto. Pero cuando se hace una búsqueda de una definición que abarque toda la complejidad de lo que implica la validez de los instrumentos de medición psicológica y educativa, lo primero que se nota es que los autores abordan cuestiones conceptuales complejas antes de ofrecer definiciones de validez. A continuación se revisan algunos de los aspectos fundamentales de la validez de los instrumentos de medición educativa, particularmente los directamente relacionados con el propósito de la validación del EXEDII.

## 4.1. Consideraciones sobre la validez de las pruebas educativas.

La validez es un concepto unitario (Messick, 1993) pero aunque con frecuencia se asume que el discurso sobre la validez de las pruebas se refiere al significado total del concepto, parecería que no es así, a juzgar por los procesos de validación que se realizan a muchos de los instrumentos de medición educativa. Es decir, es común encontrar que se sigan solamente procedimientos de validación de contenido por ejemplo, en exámenes diseñados para medir el dominio que los alumnos tienen de un contenido curricular. Por otro lado, en otros tipos de pruebas se privilegian únicamente los procedimientos de validez de criterio (concurrente o predictiva) cuando el propósito de la validación es que un examen pueda usarse en sustitución de otro asumiendo que miden lo mismo; o bien cuando se requiere predecir la ejecución de los sustentantes en una tarea, a partir de los puntajes obtenidos en la prueba. Finalmente, en un número menor de ocasiones se reportan procedimientos para evaluar la validez del constructo y, en esos casos lo que se hace es recurrir al análisis de factores.

Es importante resaltar que en el proceso de validación de un instrumento de evaluación educativa es conveniente aportar pruebas de más de un aspecto de la validez de la prueba y además, realizar un análisis lógico de los reactivos, de la adecuación, la relevancia y

representatividad de las tareas incluidas en la prueba, de las habilidades que supuestamente mide y del uso que se hará de sus resultados, así como de las consecuencias que tendrán las interpretaciones de esos resultados sobre la población evaluada (Messick, 1993). En esta forma se proporcionan elementos de la validez del constructo, que es la clase de validez que subsume a todas las demás. El problema con la validez de constructo es que muchas veces se trata de probar que una prueba mide determinado concepto que no es observable directamente y que además, es muy difícil de describir como por ejemplo la inteligencia. Por eso el proceso de la validación del constructo debe de estar sustentado por la teoría que lo explica y que define las condiciones en que éste se manifiesta.

La falta de congruencia entre la concepción de la validez y la instrumentación de los procesos de validación de instrumentos no es obra de la casualidad. El proceso histórico del desarrollo de las pruebas puede explicar el por qué de la discrepancia entre por un lado considerar que la validez es un concepto complejo y unitario y por otro aportar evidencias de validez de solamente un tipo. En los próximos párrafos se mencionan algunos de los eventos más sobresalientes del origen y evolución de las pruebas psicológicas y educativas.

El desarrollo de las ciencias naturales a lo largo del último cuarto del siglo XIX y el nacimiento de la Psicología como ciencia en 1879 dio como resultado la construcción de las pruebas psicológicas o *tests* que medían las diferencias y similitudes en la percepción de los estímulos, la velocidad de las respuestas y otros aspectos psicofísicos de la respuesta psicológica consciente de los individuos (Hilgard, 1987). Para la primera mitad del siglo XX empezaron a construirse pruebas que pretendían medir aspectos mucho más abstractos del funcionamiento psicológico. La preocupación de los psicólogos era si debían o no tratar de igual forma los puntajes emanados de todo tipo de pruebas. Es decir, las respuestas a un *test* podían interpretarse como representantes de un constructo relativamente fácil de definir, como el umbral de percepción de una señal luminosa o por el contrario, como representantes de un constructo complejo y difícil de definir como la inteligencia, los rasgos de personalidad, o el aprendizaje.

Las posturas ante este problema se dividieron y hasta se radicalizaron de manera que los puntajes de las pruebas fueron interpretados como: a) signos de procesos o estructuras

subyacentes a las respuestas, ó b) como muestras de clases de respuestas. Los autores que sostenían la primera aproximación se apoyaban en el concepto de rasgo, definido éste como una característica o un atributo estable de las personas, el cual se manifiesta en una u otra forma dependiendo del tipo de medición de que es objeto. Este concepto tiene su origen en conceptos propuestos por los filósofos griegos de la antigüedad, tales como el temperamento y el carácter, sostenidos por siglos con algunas variantes, pero que en esencia significan que las personas poseen ciertas características psicológicas internas, que prevalecen a lo largo de sus vidas. Quienes sostenían la segunda aproximación concebían las respuestas como conductas pertenecientes a una clase de respuestas, las cuales cambian en la misma forma como función de la contingencia de los estímulos. Representantes de una y otra postura son autores como Cattell y Allport, por un lado, y los psicómetras con un enfoque conductista por el otro (citados en Hilgard, 1987; Messick, 1993).

El concepto y la forma de evaluar la validez de los instrumentos fue evolucionado y sus definiciones se centraron más en sus aspectos o componentes, que en su conceptualización unitaria. En 1946, la validez de una prueba se enfocaba hacia la predicción de criterios específicos; es decir un *test* medía aquello con lo que se correlacionaba. Para 1949 Cronbach (citado en Messick, 1993) hablaba de la validez lógica, diferenciándola de la validez empírica de las pruebas. Así mismo consideraba que la validez de las mediciones de los logros (académicos) obtenidos por los estudiantes eran de otra naturaleza. La validez lógica, decía era la que se podía derivar de juicios precisos acerca de lo que mide un *test* y que se refieren a los procesos psicológicos que subyacen a las respuestas. La validez empírica se entendía como las correlaciones entre los puntajes obtenidos y alguna otra medida. Gulliksen y Lindqvist (citados en Ward, Stoker y Murray-Ward, 1996) también sugirieron que podría haber más de una validez [1]en una prueba y que se podían evaluar por separado. Este desglose de tipos de validez fue dando lugar a que en el proceso de validación de un instrumento se considerara que había algo así como un tipo ad hoc de validez que debería probarse, según el propósito de la prueba.

[1] La palabra *test*, que significa prueba en inglés, se usa en esta tesis como sinónimo de prueba o examen.

En 1954, la APA reconoció los diferentes "tipos de validez" de una prueba, a saber: validez de contenido, validez asociada a un criterio con sus dos aspectos: predictiva y concurrente y por último, la validez de constructo. Influyentes textos como los de Anastasi (edición de 1954) y Cronbach (edición de 1960) tomaron la recomendación de la APA y dividieron la validez en los cuatro tipos mencionados arriba. En las siguientes tres décadas los tipos de validez fueron tratados con diferentes énfasis, llegándose a considerar que alguno era de mayor importancia que otros.

La llamada validez de *facie*, o validez aparente fue considerada como aceptable durante la década de los cuarenta, pero dejó de serlo por no fundamentarse en métodos técnicamente adecuados para su valoración, pues dependía de la opinión no calificada de personas que con una somera revisión de los reactivos decidían si lucían válidos o no. En consecuencia autores como Anastasi (1988) advirtieron que la validez de contenido no debe ser confundida con la validez aparente ya que no se refiere a lo que la prueba mide, sino a lo que aparenta medir. Más adelante retomaremos este punto al abordar el aspecto de la autenticidad de las pruebas de idiomas, que es uno de los criterios a satisfacer para determinar la calidad de estos exámenes desde el punto de vista de autores como Bachman y Palmer (1996). Por el momento baste decir que estos autores proponen que la percepción que tiene el sustentante acerca de la autenticidad de las tareas que incluye la prueba es un factor que influye en su disposición a responderla.

Messick (1993) hace un análisis de la evolución del concepto de validez de las pruebas y tras revisar las propuestas de los autores más importantes en el tema, concluye que la validez debe entenderse como un concepto unitario y que la validez de contenido y la de criterio están implícitas en la validez del constructo. Cuando se valida el constructo se da robustez a las inferencias que se hacen a partir de los puntajes de las pruebas. Además este autor recomienda que en el proceso de validación se ponga atención a la propiedad, el significado y la utilidad de las inferencias, así como a las consecuencias sociales de la evaluación, porque eso permite tener mayor credibilidad en la generalización de los resultados de una prueba.

Para explicar su propuesta Messick hace una revisión de cada "tipo de validez" intentando rescatar lo que aporta cada uno y señalando las consecuencias de la fragmentación del concepto y los huecos que generan en la medición por medio de pruebas.

En la validación de exámenes de alto impacto la validez de contenido se establece generalmente a través de un proceso de "jueceo" que implica la comparación de cada uno de los reactivos de una prueba con la definición del contenido que supuestamente mide, expresado en términos conductuales. De esta manera la validación de contenido ofrece sustento para las interpretaciones que se hagan acerca de la relevancia y la representatividad del dominio, mas no para las inferencias que se hagan acerca de los puntajes obtenidos. Tampoco está relacionada con los procesos que subyacen a las respuestas, o con la estructura interna o externa de la prueba, ni con las diferencias de ejecución, o con las consecuencias sociales de la prueba.

La validez de criterio por su parte, está basada en la significación de las relaciones entre los puntajes de una prueba y los puntajes de otro criterio. Los resultados de correlacionar estadísticamente dos pruebas, sustentan la comparabilidad de esos instrumentos en relación con su estructura externa. Sin embargo, esa comparabilidad generalmente se refiere a la correlación entre cierto tipo de medida específica de un instrumento, con alguna otra medida particular de otra prueba en un tipo particular de circunstancia. Consecuentemente habría un número de correlaciones entre los dos instrumentos que no estarían siendo investigadas. En cuanto a la validez criterial originalmente se trataba de establecer relaciones entre los puntajes de una prueba cuyos reactivos eran muestras de conductas específicas que formaban parte de las tareas a realizar en un ambiente real. Por ejemplo, para predecir la velocidad con la que una aspirante a secretaria iba a mecanografiar un texto en la oficina, bastaba con ponerle como prueba que escribiera a máquina un texto breve y medir el tiempo que tardaba en hacerlo. Pero cuando el constructo a medir es más complejo y por tanto más difícil de definir, como es por ejemplo la vocación que tiene un alumno para estudiar Psicología, se requeriría que la prueba tuviera reactivos que implicaran la medición de actitudes, rasgos de personalidad, habilidades cognoscitivas, de hábitos y otros constructos más complejos, cuya definición suele ser imprecisa y su medición incompleta.

La conclusión de Messick respecto de "cortar" el concepto de validez para fines de facilitar el proceso de validación de las pruebas es la recomendación de considerar la validez de constructo como la integración de todas aquellas evidencias que se relacionen con las interpretaciones acerca del significado de los puntajes obtenidos en la prueba. Existen exámenes cuyo constructo se refiere a variables latentes, o a factores causales que pretenden explicar las relaciones entre sus indicadores. En esos casos el constructo es un concepto abstracto, de manera que todas aquellas pruebas que aporten evidencia acerca de que efectivamente miden distintas manifestaciones del constructo, estarán aportando sustento para la validez de constructo global. Como la validez de constructo subsume a todos los otros "tipos" de validez deberá estar respaldada, según palabras de Messick por "las evidencias y la lógica que sustentan la credibilidad de las interpretaciones de los puntajes, en términos de los conceptos explicativos que se refieren a la ejecución de la prueba y a las relaciones con otras variables" (Messick, 1993:34). Por lo tanto, el problema de cortar la validez no radica tanto en los aspectos teóricos, como en los procedimientos de validación, los cuales acusan una discrepancia entre la concepción de validez como concepto unitario y las prácticas de validación (Messick, 1988).

Habiendo llegado a este punto es posible ofrecer una definición del concepto de validez de las pruebas que incluya todos los elementos que se han considerado fundamentales en la literatura sobre el tema. La definición de Messick (Messick, 1980, 1981b, citado en Messick, 1993) establece que la validez es un juicio evaluativo global, fundamentado en evidencia empírica y en la lógica teórica acerca de la adecuación y la propiedad tanto de las inferencias, como de las acciones basadas en los puntajes de las pruebas. En esta tesis se considera que esta definición es más completa que la que se planteó al principio de este capítulo ya que abarca con amplitud el concepto de validez, refiriéndonos a todo tipo de examen, prueba o *test* que pretenda sintetizar en un puntaje, o en la calidad de una ejecución concreta y en un contexto determinado la capacidad o la habilidad, así como el conocimiento, o el aprendizaje de un individuo que responda a los estímulos de un instrumento de medición.

En los siguientes tres apartados se revisan brevemente a los tres "tipos" de validez mencionados anteriormente, con el fin de abundar en algunos aspectos relevantes a este trabajo

de investigación y fundamentar algunas decisiones que se tomaron para definir las estrategias diseñadas para lograr los objetivos que se proponía este proyecto.

### 4.2. Validez de contenido

La conceptualización de un dominio que se desea medir se puede expresar de manera tal que constituya la definición del contenido de una prueba. Por su parte, los reactivos que se elaboran para medir el manejo que los estudiantes tienen de ese dominio se consideran como una muestra de reactivos que lo representan. En ocasiones, los reactivos pueden expresarse en términos conductuales, de manera que los reactivos representarían al contenido de una muestra de conductas pertenecientes a un contenido conductual. O también pueden construirse de manera que constituyan una muestra de estímulos que, para contestarse, hagan evidente el tipo de procesos empleados o subyacentes a las respuestas. Independientemente de la manera en la que se expresen los reactivos que pretenden medir un contenido, por ejemplo el de un curso escolar se harán inferencias o predicciones a partir de las respuestas obtenidas. Entonces, la definición de un dominio que va a ser evaluado, extrayendo de él una muestra de preguntas que lo representan cabalmente en el contenido de la prueba deberá incluir la especificación de la naturaleza y de los límites del dominio, así como una apreciación de la relevancia y la representatividad de los reactivos en relación con el dominio (Messick, 1993).

La mayoría de los autores consideran que la validación del contenido de una prueba es básicamente una cuestión de juicio (Hughes, 1989: Popham, 1990; Messick, 1993 por citar algunos). Es decir cada reactivo tiene que ser valorado de acuerdo con criterios específicos para determinar su relevancia (qué tan importante es) y su representatividad (si corresponde al dominio). El proceso de valoración es realizado por jueces competentes en el manejo del contenido.

Para realizar el proceso de valoración individual de los reactivos de la prueba se requiere contar con una definición adecuada. Así, dependiendo del propósito y la naturaleza de la prueba, la definición del dominio debe incluir aspectos que garanticen a) que el contenido de

los reactivos sea una muestra del contenido del dominio; b) que las conductas que se realicen en la resolución de la prueba sean una muestra de las conductas que se emitirían en una ejecución típica y c) que los procesos empleados por el sustentante para llegar a las respuestas sean los que típicamente estarían involucrados.

De acuerdo con lo anterior, una prueba de certificación, por ejemplo, deberá definir los niveles mínimos deseables de conocimientos y habilidades adquiridos, asociados a un nivel determinado de manejo o maestría. Por tanto la definición del contenido deberá estar expresada en términos de: a) los objetivos de un contenido académico; b) las conductas que típicamente se emiten en la manifestación del manejo de ese dominio; y c) la definición de las habilidades y conocimientos adquiridos que se relacionen plausiblemente con el dominio (Messick, 1993).

### 4.3. Validez de criterio

La validez criterial se evalúa comparando los puntajes de una prueba con una o más variables externas (llamadas criterios), que se considera que proveen una medida directa de la característica o conducta en cuestión. Para sustentar este tipo de validez se recurre usualmente a realizar análisis estadísticos de correlación o regresión entre los puntajes de la prueba y los puntajes del criterio. Es decir, los puntajes de la prueba que se quiere validar se comparan con los puntajes obtenidos en otra prueba, o con otro criterio como pueden ser las calificaciones en ciertas materias escolares. Dado que la validez criterial se apoya sólo en ciertas partes de la estructura externa del examen, la comparación de los puntajes de la prueba con el criterio no resulta en el establecimiento de, por ejemplo patrones de relaciones de los puntajes con otras medidas, sino que se encamina a ciertas relaciones entre medidas específicas, las cuales son como se dijo antes, criterios para un propósito particular aplicado, en un ambiente específico.

La validez criterial se divide en: a) validez predictiva la cual se identifica con el propósito de predecir la ejecución de los sustentantes en una tarea, dentro de un escenario auténtico a partir de los puntajes obtenidos en la prueba. Es decir indica el grado en que la prueba predice un criterio en cuestión; b) validez concurrente la cual indica el alcance con que el puntaje de la

prueba mide el constructo, en función de su correlación con los puntajes obtenidos por otra medida similar, como puede ser otro instrumento ya validado y consolidado. En términos de la temporalidad de la aplicación de las pruebas, si el criterio es obtenido cierto tiempo después de que la prueba se ha administrado se habla de validez predictiva. Si por el contrario, el criterio se determina al mismo tiempo, se trata de validez concurrente (Hernández Sampieri, Fernández Collado y Baptista, 2003).

### 4.4. Validez de constructo

La validez de constructo incluye a todos los demás "tipos" de validez. Por lo tanto, aportar evidencias acerca de la validez del contenido constituye un respaldo para las inferencias que se hagan con respecto a la relación que existe entre el dominio definido y los reactivos de la prueba. Por su parte las evidencias de validez de criterio, sean de tipo concurrente o predictiva apoyan las inferencias que se hagan acerca de las relaciones que hay entre el dominio que miden dos pruebas independientes en el primer caso y las relaciones entre una prueba y la ejecución de las conductas que define el dominio pasado un lapso determinado, en el segundo caso. Por lo tanto, cuando los estudios aportan evidencias de validez de contenido y de criterio se están aportando evidencias para la validez del constructo. No obstante, abordar la validez de constructo implica más que la acumulación de evidencias de otros tipos de validez ya que la validez de constructo sustenta las inferencias que se derivan de los puntajes de las pruebas.

Las evidencias de validez de contenido pueden aportar peso a la validez de constructo en el grado en el que el contenido sea representativo y relevante. La relevancia del contenido de una prueba se puede derivar, según Messick (1993) de tres fuentes: a) juicios de profesionales acerca de la relevancia y representatividad del contenido en relación con el dominio; b) evidencia de que los resultados reflejan procesos que se consideran importantes en el dominio de la función y c) correlaciones significativas con medidas criteriales de la ejecución del dominio.

Con base en lo anterior pueden existir dos tipos de problemas en la interpretación de la validez de una prueba. El primero es que dado que la mayoría de los exámenes no pueden

incluir el contenido completo del constructo en la prueba, se evalúa solamente una muestra de ellos suponiendo que representan todo el dominio. Es muy importante asegurar que la muestra de contenidos evaluados sea representativa, ya que si una prueba deja de medir aspectos importantes del dominio por no estar incluidos en su contenido, la prueba tendrá una representatividad escasa (Messick, 1993).

El segundo aspecto de importancia en la consideración del contenido de un examen es su relevancia. Es decir, si el contenido de un examen incluye factores que no son relevantes para representar al dominio, entonces las interpretaciones que se hagan de los puntajes obtenidos estarán contaminadas con aspectos que no corresponden al dominio. Kenyon (1998, citado en Hui Ling, 2003) dice que los juicios acerca de la relevancia de las tareas de una prueba son un aspecto fundamental de la representatividad del contenido del examen. Hui Ling (2003) recomienda que para asegurar la relevancia del contenido de una prueba es necesario tomar en cuenta todos los aspectos de la situación de prueba, incluyendo la especificación del dominio del constructo, en términos del conocimiento que se espera que el sustentante tenga, con respecto al tema, las especificaciones de la prueba, los reactivos, las condiciones de administración del examen y los criterios para su calificación.

Recapitulando lo dicho hasta ahora: 1) existen dos enfoques desde los que se pueden interpretar los resultados que los sustentantes obtienen en los exámenes: a) siguiendo la metodología normativa que implica comparar los resultados obtenidos en la prueba con los resultados de otros sustentantes que han contestado el mismo instrumento en condiciones similares, con el objeto de obtener una visión general de la variabilidad del aprovechamiento académico de los educandos en una población dada; o b) siguiendo la metodología criterial, comparándolos con las descripciones de contenidos curriculares específicos para verificar la maestría con la que cada educando maneja un conjunto de contenidos académicos dentro de un dominio determinado. 2) El origen, desarrollo y características de los exámenes estandarizados han dado como resultado su uso generalizado para sustentar interpretaciones que llevan a la toma de decisiones en el ámbito de la educación. Algunas de esas decisiones suelen tener consecuencias de alto impacto en los sustentantes, en los usuarios y en otros sectores sociales. 3) De las características que le confieren credibilidad a un examen la más importante es sin duda

la validez. La evolución del concepto de validez y las maneras en las que se ha intentado operacionalizar el concepto en las pruebas psicológicas y educativas acusan algunas discrepancias entre lo que la teoría estipula y recomienda y lo que las pruebas como instrumentos de medición realmente miden.

En el siguiente capítulo se abordan algunos puntos de la evaluación de uno de los constructos más complejos y más difíciles de medir, como es el lenguaje. La aproximación a ese tema no parte del enfoque propio de la lingüística, sino del de la evaluación psicométrica del dominio de la lengua, enriquecido para propósito de esta investigación, con un modelo que propone la evaluación de elementos que permiten un análisis más extenso y al mismo tiempo más profundo de las evidencias de la validez de una prueba que mide el lenguaje. Así, expone un marco conceptual que fundamenta la metodología seguida en la investigación y la manera en la que se interpretan los hallazgos obtenidos.

### 4.5. La validación de una prueba del manejo de un idioma.

Partiendo de la concepción del lenguaje como una característica definitoria del ser humano que le permite comunicarse con su entorno podemos dar una definición sencilla, que incluya los elementos esenciales de ésta. El diccionario de la Real Academia Española (2001) dice que el lenguaje es el conjunto de sonidos articulados con los que el hombre manifiesta lo que piensa o siente. La lengua por su parte, se define como un sistema de comunicación y expresión verbal propio de un pueblo o nación, o común a varios. El hecho de que se entienda la lengua como un sistema implica que los elementos que lo conforman y que se usan para comunicar y expresar se apegan a reglas para combinar sus componentes. Es de notarse que los términos mencionados en las dos definiciones hacen referencia a los pensamientos y a los sentimientos de las personas, constituyendo éstos dos constructos que como el lenguaje, son extremadamente difíciles de definir y de medir.

La aproximación estructural del estudio del lenguaje que había estado vigente durante casi todo el siglo XX sufrió cambios a partir de los primeros años de la década de los setenta con

la propuesta de Chomsky quien, junto con Saussure contribuyó a la consolidación de la lingüística como ciencia en el siglo XX (Santos Caicedo, 2007). Chomsky propuso una elaboración más compleja de la estructura gramatical que supone

> ...un sistema de reglas finito que genera una pluralidad infinita de estructuras profundas y superficiales, adecuadamente relacionadas entre sí. Debe también contener determinadas reglas que establezcan la relación entre esas estructuras abstractas y ciertas representaciones del sonido y el sentido, representaciones que es de presumir que están constituidas por elementos pertenecientes a la fonética universal y a la semántica universal. (Chomsky, 1971, p. 35; citado en Santos Caicedo, 2007).

Los conceptos de competencia y ejecución (*performance*) de la teoría de Chomsky abrieron la perspectiva social del lenguaje a la lingüística, pero fue Hymes (1971) quien cuestionando a Chomsky desarrolló el concepto de competencia comunicativa, la cual supone el conocimiento del uso contextual o sociolingüístico del lenguaje. Más tarde, los trabajos de Halliday (1982) aportaron los conceptos de nociones y funciones del lenguaje, lo que permitió poner énfasis en las funciones del lenguaje y en la adquisición del mismo. Sus trabajos constituyen el sustento teórico de los procesos semióticos del lenguaje. Por otra parte, los trabajos de Widdowson (1978) y de Brumfit y Johnson (1979) y posteriormente los de Canale y Swain (1980) tuvieron un fuerte impacto en la investigación y el desarrollo de la lingüística aplicada y a la enseñanza de las segundas lenguas y las extranjeras, campo en el que se desarrolla la propuesta de Bachman (1990) y el modelo de Bachman y Palmer (1996). La integración de las propuestas de estos y algunos otros autores generó otro enfoque llamado comunicativo, que considera al usuario de la lengua como un protagonista con sus necesidades de comunicación en las distintas situaciones en que tienen lugar.

En el ámbito de la Psicología el estudio del lenguaje ha sido abordado desde diferentes puntos de vista, desde la concepción del lenguaje como una conducta sujeta al condicionamiento operante (Skinner, 1957/1981), hasta concebirlo como una manifestación del pensamiento (Piaget, 1984), o bien como herramienta que el ser humano utiliza no solamente para comunicarse con sus congéneres, sino que le permite construir la noción del mundo que le rodea, de acuerdo a la postura de Vygotsky (citado en Hernández, 2005). Es importante notar la

diversidad de planteamientos teóricos sobre el origen, desarrollo y función del lenguaje ya que de una u otra forma han influido en la enseñanza y en la evaluación de los idiomas y por ende, en la construcción de los exámenes como instrumentos de medición de habilidades, destrezas y competencias o dominios.

En otro contexto el lenguaje puede constituir una materia curricular que se enseña en las escuelas y que por lo tanto puede ser evaluada a través de exámenes u otro tipo de pruebas. El manejo de la lengua nativa u oficial, como es el español en los países hispanoparlantes, o el inglés en los países anglosajones se evalúa comúnmente a lo largo de la instrucción básica, con énfasis en el conocimiento de la gramática, por lo que los exámenes suelen contener reactivos que prueben el manejo de reglas y funciones de las palabras en la oración. Pero la evaluación no debe concretarse a la parte estructural de la lengua; es necesario abordar aspectos cognoscitivos y sociolingüísticos; afectivos y operativos, lo mismo en el caso del aprendizaje de la lengua materna, como en el de una segunda lengua, o de lenguas extranjeras.

Se considera que hay diferencias importantes entre el aprendizaje y uso de una segunda lengua y el aprendizaje y uso de una lengua extranjera. DiMatteo (2004) ubica en el caso del inglés, seis diferencias esenciales entre dos situaciones en las que los estudiantes aprenden una lengua distinta de la lengua materna. Lo mismo puede decirse sin duda de otros idiomas, además del inglés pero este trabajo se aboca a ese idioma por ser el que mide el EXEDII. En la Tabla 4.1 se resumen las diferencias entre segunda lengua y lengua extranjera, de acuerdo con DiMatteo (2004).

Tabla 4.1. Diferencias entre el Inglés como segunda lengua y como lengua extranjera (EFL y ESL, por sus siglas en Inglés). Tomado de DiMateo, 2004.

| Inglés como segunda lengua (ESL) | Inglés como lengua extranjera (EFL) |
|---|---|
| El estudiante vive inmerso en un ambiente en el que se habla Inglés | El estudiante vive inmerso en un ambiente en el que se habla su lengua materna |
| La práctica, el reforzamiento y la confirmación de la lengua tienen lugar en un ambiente fuera del salón de clases. Los estudiantes ganan confianza por ellos mismos. | El único ambiente en el que se habla inglés es el que el estudiante encuentra en el salón de clases y es el profesor el que proporciona la práctica, el reforzamiento y la confirmación para que el alumno gane confianza en la lengua. |
| Los estudiantes aprenden la gramática del inglés al mismo tiempo que están inmersos en un ambiente angloparlante. | Los estudiantes solamente estudian la gramática inglesa no coloquial, a menos que estén inscritos en un nivel avanzado. |
| Las clases pueden tener más de 20 horas de duración en una semana. | Las clases de inglés pueden abarcar de una a cinco horas por semana. |
| Se requiere que el alumno asista presencialmente a cada clase. | Las clases se programan entre otras actividades del estudiante. No son prioritarias, como otras materias. |
| Los grupos son de 12 a 15 alumnos, siendo la clase centrada en el maestro. | Los grupos son de 1 a 8 alumnos, siendo la clase centrada en el alumno. |
| El maestro define el currículo | A menudo se identifican las necesidades del alumno (ayudarlos a hablar sobre sus intereses académicos, por ejemplo) y se les va instruyendo poco a poco. |
| Las expectativas y la motivación del maestro y de los alumnos son altas. | Las expectativas y la motivación del maestro y de los alumnos son bajas. |

El punto medular de la diferencia entre el manejo del inglés como segunda lengua (ESL) y el manejo del inglés como lengua extranjera (EFL ambos por sus siglas en inglés) es el grado de inmersión en el que el educando aprende la lengua. El estudiante que aprende una lengua distinta de su lengua materna, que lo escucha, lo lee y lo usa en su vida cotidiana tiene una inmersión total y termina por integrarlo como una segunda lengua, es decir llega a ser bilingüe. En cambio, el estudiante que solamente le dedica unas horas a la semana y que sus únicas oportunidades de escucharlo, hablarlo y practicarlo son las que tiene en el ambiente de la clase

tiene una inmersión parcial y tarda más tiempo en llegar a ser bilingüe, si es que lo logra. Este último es el contexto en el que se desarrollan la mayoría de los estudiantes de la UABC, que son los sustentantes habituales del EXEDII.

Las implicaciones que puede tener la adquisición del inglés como segunda lengua, o como lengua extranjera son diversas. Entre otras es claro que la inmersión total en el idioma también implica una inmersión en las vivencias características del ambiente, de tal manera que se aprenden aspectos sutiles del uso del lenguaje que no pueden ser incluidos en el currículo escolar. Por ejemplo en el uso de la lengua materna, las personas fácilmente pueden detectar el registro, o la clase de lenguaje que debe usarse dependiendo de la situación. Es decir, no se usa el mismo lenguaje en una situación formal, que en una informal. El uso de expresiones, posturas corporales o actitudes son parte de la expresión del mensaje. Por estas razones los exámenes del dominio de un idioma deberían explicitar en sus objetivos qué aspectos del manejo de la lengua pretenden medir y diseñar formas adecuadas para hacerlo (Bachman y Palmer, 1996).

### 4.5.1. Evaluación del inglés como segunda lengua.

El idioma Inglés es hoy una lengua que permite la comunicación entre más de 300 millones de personas que lo hablan como lengua materna, así como varios millones más que lo usan como segunda lengua. También hay un número elevado de personas que lo han aprendido como una lengua extranjera con fines de comunicación en el amplio mundo del aprendizaje escolar, el turismo, la diplomacia, la ciencia, la política y otros contextos. Una de cada cinco personas en el mundo hablan Inglés con un nivel de competencia aceptable y se espera que en un futuro cercano el número de personas que hablen inglés como segunda lengua será mayor que el número de personas que lo hablan como lengua nativa (*AskOxford*, 2005).

Por otra parte, el fenómeno de la migración en todo el mundo y particularmente el que se da en da entre México y EUA ha hecho que el aprendizaje del idioma inglés sea una necesidad urgente para miles de personas. El número de ciudadanos mexicanos que cruzan la frontera mexico-estadounidense, con intenciones de encontrar trabajo en aquel país se duplicó entre

1990 y 2000, aumentando de 2.6 a 4.9 millones de personas, de acuerdo con el Censo de ese último año (*US Census Bureau's Census 2000 Public Use Micro-Sample*, citado en *Migration Policy Institute*, 2004). En el año 2002 el país de origen del mayor número de inmigrantes legales a ese país fue México con 219,380, aproximadamente el 20.6% del flujo total (INS, Anuario Estadístico 2002 del Servicio de Inmigración y Naturalización, Año Fiscal 2000-02, citado en el sitio Web de la Embajada de los Estados Unidos en México, 2004). Además, la población indocumentada de México que reside en los Estados Unidos se incrementó de cerca de 2 millones en 1990 a 4.8 millones en el año 2000. (*Census Bureau 2003. "Estimates of the unauthorized immigration population residing in the U.S.*"). Sin ánimo de poner adjetivos a la situación económica y política que prevalece se señala la dirección en la que probablemente se moverá el desarrollo de los países en los próximos años y México no podría moverse en otra dirección, sin verse afectado por ello.

Es un hecho que la migración de grandes grupos de personas que buscan mejores oportunidades de vida en los países más desarrollados del planeta provoca cambios en las políticas internas, en las economías locales y en la transformación cultural de los sitios en los que se concentran poblaciones de inmigrantes. Como se dijo arriba, la población hispanoparlante, aunada a la población asiática y de otros países que se han establecido recientemente en los EUA ha impactado el ámbito de las escuelas, obligando al gobierno a tomar medidas para asimilar a estos ciudadanos --legales o no-- que acuden a los centros escolares, de salud y otros servicios. La ley estadounidense denominada *No Child Left Behind* (Ningún niño dejado atrás) vigente desde 2002 establece que las escuelas primarias estadounidenses atiendan las necesidades particulares de millones de niños que tienen algún problema para funcionar como alumnos regulares. El caso de los niños que no tienen el inglés como primera lengua, de los cuales el 76% corresponde a los niños de origen hispano (citado en Abedi, 2003) hace más urgente que todos esos estudiantes tengan que aprender el inglés para poder funcionar como alumnos regulares.

Esta circunstancia ha hecho que se vuelva a revisar el tema de las implicaciones de la evaluación de las lenguas en los EUA y en otros países que reciben grandes cantidades de inmigrantes. Diferentes estados de ese país han tomado medidas específicas para disminuir el

impacto negativo que ejerce el hecho de no ser angloparlante nativo sobre los resultados de la evaluación de alto impacto. Tales medidas van desde exentar de esas evaluaciones a los estudiantes con limitado manejo en el Inglés (LEP por sus siglas en Inglés) hasta permitirles más tiempo para completar sus exámenes, tener explicaciones adicionales, o presentar los exámenes en grupos pequeños, entre otras (Holmes y Duron, 2000). En esencia lo que se trata de lograr con todos esos esfuerzos es asegurar que los instrumentos y formas de medición que determinan aspectos importantes de la vida de los estudiantes presenten el menor sesgo posible, es decir que sean válidos para todas las poblaciones evaluadas.

Hablar de validez en los exámenes que evalúan el manejo de lenguas extranjeras, o de una segunda lengua es un tema que ha merecido la atención de investigadores que toman una de las dos siguientes posturas: a) el lenguaje es un proceso complejo que comparte habilidades que intervienen en los procesos cognoscitivos que tradicionalmente se engloban en el concepto de inteligencia (Oller, 1979) y b) el lenguaje es un conjunto de habilidades que se manifiestan de diversas maneras en el proceso de comunicación, entendido éste como una experiencia creativa de interacción entre personas (Bachman, 1990; Alderson, 1993). Este último es el enfoque desde el que se realiza esta investigación.

### 4.5.2. Evaluación del inglés como lengua extranjera

El propósito de los exámenes de idiomas es el de hacer inferencias acerca de la habilidad que los sustentantes tienen en una lengua específica distinta de su lengua materna. Con frecuencia, esas inferencias son la base para tomar decisiones acerca de éstos, como la ubicación en un nivel determinado de instrucción, o la asignación a un grupo remedial; el ingreso o egreso de una institución educativa, o la obtención de un puesto laboral, por poner algunos ejemplos. Decisiones como éstas pueden tener consecuencias importantes en la vida de las personas por lo que resulta primordial hacer todos los esfuerzos posibles por lograr que los instrumentos de medición del manejo de lenguas sean confiables y válidos.

La confiabilidad y la validez pueden construirse junto con la prueba misma si en el proceso de construcción se cuidan escrupulosamente los elementos que determinan estos dos factores. Como se menciona en el apartado 2.3 en los exámenes normativos se efectúan análisis estadísticos con los datos de las aplicaciones piloto y aplicaciones subsecuentes con objeto de confirmar o afinar aspectos como la confiabilidad, el nivel de dificultad, la discriminación y el funcionamiento de los reactivos. En los exámenes criteriales pueden hacerse esos mismos análisis pero en general la calidad de la prueba depende más del método utilizado para su construcción y de las decisiones tomadas por los comités de jueces respecto de la definición de los contenidos, el tipo y número de reactivos, las características de las opciones de respuesta, el punto de corte y todos los aspectos que tengan que ver con las características de la prueba y su capacidad para medir lo que pretende medir.

En los procesos de validación del contenido de un examen, sea éste criterial o normativo se precisa de la discusión y las decisiones consensuadas de grupos de expertos que trabajan en torno a un propósito y de una manera sistemática, ajustándose a criterios y haciendo todo esfuerzo posible por mantener una deseable objetividad.

Es claro que cualquier actividad humana tiene una fuerte carga subjetiva y que la elaboración de juicios, aunque sean hechos por expertos en una materia, sigue siendo una actividad humana y por tanto subjetiva. Sin embargo hay autores como HajiPourNezhad (2003) quien propone nueve principios a los que es necesario apegarse para aprovechar las ventajas del procedimiento llamado Moderación o jueceo el cual consiste en la reunión de expertos en evaluación para revisar, discutir y evaluar los materiales de una prueba. Los principios a los que este autor propone apegarse son los siguientes:

1) Los juicios son inevitables al construir o validar pruebas de desempeño en lenguas.
2) Ningún experto en evaluación puede hacer juicios perfectos, pero los juicios realizados en grupo (con ciertas limitaciones) pueden acercarse más a la perfección, que los juicios individuales.

3) Los juicios logrados en comités de jueces expertos en evaluación deben ser expresados en términos cuantificables para favorecer su precisión.

4) Los juicios logrados en comités de jueces expertos, independientemente de la cuestión que debatan, deben incluir opiniones acerca de las opciones alternativas. Y los consensos deben ser alcanzados con base en medidas estadísticas claras.

5) La validación a posteriori de una prueba no puede revertir todos los juicios (subjetivos) importantes hechos en el proceso de la construcción. Pero muchos juicios sí pueden influir en la forma en la que una prueba fue construida y en la manera en que afecta la ejecución de los sustentantes sin quedar atrapados en la validación retrospectiva de la prueba.

6) Los juicios logrados en un comité de expertos debidamente informados, deben ser utilizados en todos los aspectos de la prueba y no solamente en los comités que se conforman para decidir las tareas de la prueba y el nivel de habilidad del candidato.

7) Dado que las tareas de la prueba deben reflejar las necesidades de la situación meta, es necesario que los comités de expertos debatan desde el inicio de la construcción de una prueba, enfocándose en la identificación de las necesidades de los niveles teóricos y prácticos. Esto involucra también la identificación de las suposiciones de los expertos acerca de cada situación particular de prueba. Si estas suposiciones no son investigadas, influenciarán el proceso de la construcción de la prueba sin ser notadas.

8) Las entrevistas individuales para colectar las opiniones o juicios de los expertos, acerca de determinados aspectos de las pruebas deben ser sustituidos por la formación de comités en los que los jueces estén reunidos y tengan oportunidad de discutir y llegar a juicios consensuados.

9) Durante el proceso de debate de los jueces, en la evaluación de una prueba a posteriori, debe contarse con un listado de los aspectos de la prueba que van a ser evaluados por los jueces incluyendo las afirmaciones que son consideradas como correctas, incorrectas o dudosas, claramente indicadas y los jueces deberán utilizar estos listados para la evaluación del conocimiento y de los

juicios de los constructores de la prueba. Estas afirmaciones pueden ser colectadas de los mismos expertos que participaron en la prueba, quienes funcionarán como informantes del estudio y/o del propio investigador.

Se hace mención de lo anterior por tres razones: a) los exámenes criteriales como el EXEDII se construyen y validan utilizando paneles de jueces para tomar las decisiones que determinan la calidad de la prueba, b) en la validación del contenido se recurrió a paneles de jueces y c) el modelo de Bachman y Palmer el cual proporciona el enfoque de este estudio de validación emplea el método de moderación o jueceo.

### 4.5.3. El modelo de Bachman y Palmer

Lyle F. Bachman y Adrian S. Palmer (1996) propusieron un marco conceptual en el que se pretenden incluir todos los aspectos que la literatura señala como relevantes para la calidad de un instrumento, siendo el indicador de calidad más importante la validez. Las pruebas no son confiables o válidas; lo que se valida son ciertas interpretaciones que se hacen de los puntajes obtenidos en ellas. Es tan amplio y tan variado el universo de factores que pueden influir en los resultados de un examen que no se podría validar su uso con todo tipo de poblaciones, o generalizar las interpretaciones sobre los puntajes obtenidos a otros constructos aparentemente similares, o para poblaciones con distintas características. Para poder confiar en que un examen mide lo que pretende medir es necesario, si no controlar todas las variables relacionadas con la evaluación, sí por lo menos reconocerlas y tratar de minimizar el impacto de no tomarlas en cuenta.

En el enfoque evaluativo tradicional se concibe a la educación escolar dentro de una organización en niveles o grados, con un currículo elaborado en términos de metas a alcanzar por los alumnos a través de la guía de sus profesores. Este enfoque se deriva de una clasificación de las metas educacionales por objetivos, los cuales son definidos en términos de las conductas que los estudiantes deberán emitir para demostrar que han adquirido los conocimientos planeados para cada nivel. La evaluación del logro de esos objetivos implica el

diseño de instrumentos que midan el grado de logro que cada estudiante ha tenido en los objetivos del currículo. El método de clasificación y redacción de objetivos que utiliza el enfoque tradicional es el que propusieron un grupo de psicólogos reunidos primeramente en una convención de la APA en Boston, EUA, en 1948 y en numerosas ocasiones posteriores. Este grupo buscaba una solución a los problemas de comunicación que prevalecían entre los educadores en general, quienes no llegaban a acuerdos sobre la mejor forma de enseñar y de evaluar en las escuelas debido a que no existía una forma objetiva y estandarizada de definir y de comunicar los productos deseables de la instrucción escolar. Como resultado de estas reuniones se escribió el libro Taxonomía de los objetivos educativos (Bloom, 1956, 1984) que proporcionó un marco de referencia en el que se construyeron programas académicos y exámenes y que ha prevalecido hasta estos tiempos, con pocos cambios de fondo.

Pero en esa convención de 1948 se había planeado fundamentar y elaborar las taxonomías de tres dominios: el cognoscitivo, el afectivo y el psicomotor. La taxonomía del dominio cognoscitivo fue completamente desarrollada, pero las otras dos sólo se trabajaron parcialmente. En palabras de Bloom:

> [...] es difícil describir las conductas apropiadas en relación con estos objetivos [afectivos], desde el momento en que los sentimientos y emociones interiores son tan significativos en su dominio, como las manifestaciones de una conducta determinada. Al mismo tiempo además nuestros métodos de comprobación y examen en el campo afectivo no han logrado todavía superar las fases más elementales..." (Bloom, 1981, p. 8-9).

Existe acuerdo entre los expertos en la construcción y uso de exámenes educativos en cuanto a la necesidad de operacionalizar las variables que intervienen en el proceso enseñanza-aprendizaje-evaluación, tomando en cuenta factores como el contexto de la aplicación de las pruebas y las características de los sustentantes, así como las consecuencias del proceso de evaluación. No obstante, las pruebas que se han desarrollado hasta hoy no cumplen cabalmente con esta recomendación debido a las dificultades teórico-prácticas de integrar dentro de un mismo instrumento de evaluación todos los conceptos que se consideran relevantes para garantizar la validez de las interpretaciones derivadas de los puntajes obtenidos en ellos. Es por

eso que el modelo de Bachman y Palmer se considera valioso en este trabajo de tesis, puesto que proporciona una conceptualización de la medición del manejo de un idioma extranjero en el que se reconoce la influencia de numerosos factores que aportan algo al proceso y que en ocasiones parecen tan obvios que se dan por sentados, como las condiciones de iluminación de la sala de evaluación, o la motivación de los sustentantes. Existen otros factores tal vez menos obvios que no solamente han sido tomados en cuenta, sino que han sido objeto de teorías acerca del lenguaje como son los aspectos genéticos, o neurofisiológicos. Pero independientemente de que los procedimientos de construcción o validación de exámenes para evaluar el lenguaje expliciten, definan o controlen mayor o menor número de variables lo importante es tener la capacidad de detectar, definir y medir las variables más relevantes en cada caso en particular. En este sentido, dependiendo de la naturaleza del instrumento de medición de que se trate, así como del propósito del estudio pueden considerarse más relevantes algunas variables, sin que se ignoren completamente otras.

A continuación se explican algunos de los conceptos fundamentales del modelo de Bachman y Palmer con el objeto de ilustrar los aspectos del modelo que se tomaron en cuenta para esta investigación.

#### 4.5.3.1. El concepto de uso del lenguaje.

En lenguaje coloquial nos referimos al dominio de un idioma en términos de "hablar" esa lengua. Es claro que cuando decimos que alguien "habla Inglés" estamos implicando que también lo entiende, lo lee y lo escribe con alguna destreza, no necesariamente equivalente en los cuatro aspectos. Durante varias décadas se ha considerado que existen cuatro habilidades básicas en el manejo de un idioma: dos habilidades receptivas, que son escuchar y leer y dos habilidades productivas, que son hablar y escribir. Pero esta idea y esta forma de clasificar las habilidades subyacentes al manejo de una lengua son insuficientes para cubrir todas las áreas que abarca el lenguaje en la vida de una persona. La idea de usar un idioma es más cercana a la idea de usar una herramienta, por ejemplo.

El uso que un individuo hace del lenguaje para comunicar todo lo que tiene que comunicar en el ambiente natural adquiere características específicas dependiendo de la situación en la que se da. El uso del lenguaje en una ponencia en un congreso, por ejemplo es distinto al de la plática entre amigos. De acuerdo con Bachman y Palmer (1996) cada una de estas circunstancias tiene su propio dominio: el Dominio de Uso del Lenguaje (*Language Use Domain*) . Para efectos de medición es posible descomponerlo en tareas específicas del uso del lenguaje según una muestra dada, es decir el Dominio del Uso de la Lengua Meta (DULM) por ser el dominio que se va a evaluar. Este dominio es más amplio que el dominio que cubren los reactivos de la prueba en términos de las tareas que la conforman, pero si las tareas del examen son relevantes y pertinentes representarán al dominio DUML y por lo tanto las interpretaciones que se hagan a partir de los resultados de la prueba, podrán generalizarse a ese dominio.

Entonces, el DULM es el conjunto de tareas específicas al uso de la lengua que el sustentante probablemente encontraría fuera de la situación de prueba y al cual se desean generalizar las inferencias acerca de la habilidad que el sustentante tiene en el idioma. Dicho en otras palabras la habilidad en el idioma se conceptualiza como el conjunto de tareas que el individuo efectúa cuando usa el lenguaje, en una situación determinada. Por eso una prueba para evaluar la habilidad del sustentante en un idioma debe contener un número de tareas a resolver, que tengan las características que las hacen relevantes y representativas de ese dominio, es decir el DULM. De aquí se puede inferir que una prueba tendrá validez de contenido cuando las tareas que se incluyen como reactivos de la prueba sean esencialmente iguales a las del DULM.

De lo anterior se deriva que para evaluar el contenido de una prueba, primero es necesario hacer una descripción de las tareas del DULM y luego cotejarlo con los reactivos para ver si se corresponden mutuamente. En este sentido es posible definir un examen de la habilidad del lenguaje como un procedimiento para provocar oportunidades del uso de la lengua, a partir del cual se puedan hacer inferencias acerca de la habilidad que el sustentante tiene en ese idioma.

### 4.5.3.2. Las características de las tareas del DULM y de las tareas de la prueba.

Las tareas del DULM tienen características que las definen como miembros de ese dominio. Con la ayuda del marco conceptual propuesto por Bachman y Palmer (1996) es posible identificar, definir y clasificar las características del DULM para construir o evaluar una prueba del manejo de la lengua, verificando que las tareas sean esencialmente iguales a las que realizaría el sustentante en un ambiente distinto del de la prueba. Estas características son: características de la administración, de la rúbrica, del estímulo, de la respuesta esperada y de la relación entre estímulo y respuesta. Cada una de ellas tiene varios aspectos a considerar, los que se explican en los siguientes párrafos.

- Características de la administración. Son las circunstancias físicas en las que puede ocurrir el uso de la lengua o la prueba. Son de tres tipos:
    - a) físicas como el lugar, el nivel de ruido, la temperatura, la iluminación, el tipo de asientos, el equipo con el que el sustentante tiene que trabajar, etc.
    - b) dos tipos de participantes: los sustentantes y los encargados de administrar el examen; también se refiere a los roles que éstos juegan en la situación, como por ejemplo en una entrevista, o cuando el sustentante resuelve el examen y el administrador sólo cuida que todo esté en orden, etc.
    - c) momento de la aplicación, que hace referencia a si el sustentante está fatigado, o fresco y dispuesto a la tarea.

**Características de la rúbrica**. Se refiere a la mecánica para resolver la prueba y la forma en la que va a ser calificada. Tiene cuatro aspectos:

- la estructura de la prueba, o sea la forma en la que está organizada. Esta a su vez consta de cuatro tipos de características:
    - i) número de partes, que se refiere a las partes de que consta la prueba, o al número de tareas;

- ii) diferenciación de las partes que significa que cada tarea es claramente distinguible de otras;
- iii) secuencia de las partes que puede ser fija o variable
- iv) importancia relativa de las partes que se refiere a que algunas partes de la prueba pueden ser más importantes que otras;
- v) número de reactivos por cada parte de la prueba;

- b) instrucciones las cuales deben de ser explícitas por la necesidad de hacer inferencias acerca de la ejecución del sustentante y debe siempre informársele acerca de cómo responder al examen, cómo va a ser calificado y cómo van a ser usados sus resultados. Las instrucciones pueden llegar al sustentante de dos maneras:
  - i) lenguaje que puede ser el nativo o el evaluado y
  - ii) a través de un canal que puede ser aural, visual, o ambos.
- c) tiempo disponible para resolver la prueba el cual puede ser
  - i) cronometrado o
  - ii) libre;
- d) método para calificar que se refiere a la forma en la que a las respuestas del sustentante se le van a asignar números o calificaciones. Hay tres maneras de hacerlo:
  - i) criterio de corrección que es cuando existe una clave que indica cuáles respuestas son correctas o cuando se usarán escalas variables o bien, se efectuarán juicios;
  - ii) procedimientos para calificar las respuestas que son los pasos a seguir en la calificación es decir si son secuenciados o no; si son calificados por los mismos jueces, o por diferentes;
  - iii) explicitación de los criterios y procedimientos es el grado en el que los sustentantes son informados de la forma en la que se les va a calificar es decir de la naturaleza de los criterios de calificación; si los resultados van a ser entregados u omitidos o bien si se deja este punto deliberadamente vago para el sustentante.

- **Características del estímulo**. Es el material que contiene una prueba determinada, o de una tarea de la prueba, la cual se espera que el sustentante procese en alguna manera y para las que se espera una respuesta. Se describe en términos de las características del formato y del lenguaje:
- a) formato es la forma en la que se presenta el estímulo en la prueba e incluye las siguientes características:
    - i) según el canal que puede ser aural, visual, o ambos;
    - ii) según la forma que puede ser usando lenguaje o no, es decir con figuras o dibujos o ambos;
    - iii) tipo de lenguaje que se usa es decir el nativo, el que va a ser evaluado, o ambos;
    - iv) tamaño siendo éste en palabras, frases cortas, oraciones, párrafos, o discurso. El tamaño del estímulo tiene influencia en la cantidad de interpretación que va a ser requerida por el sustentante. Estímulos cortos en general implican menos interpretación que los largos;
    - v) tipo de estímulo que puede ser en forma de un reactivo, o de un comando o *prompt*. Un reactivo es un segmento de lenguaje/no lenguaje altamente enfocado a un determinado fin; su finalidad es provocar una respuesta de selección o de respuesta corta. En el uso del lenguaje en el medio natural, corresponde a estímulos cortos como los que se dan en una conversación telefónica y que provocan respuestas cortas como por ejemplo: "sí", "¿en serio?", "mhmm". Un *prompt* es un estímulo que constituye una orden o comando que produce una respuesta extensa como "redacta una composición";
    - vi) grado de velocidad es la rapidez con la que el sustentante tiene que procesar la información del estímulo;
    - vii) vehículo es el medio por el que se entrega el estímulo que ser en vivo, reproducido o ambos;
- b) lenguaje del estímulo en las tareas en las que el estímulo se da usando lenguaje; éstas características se refieren a la naturaleza del lenguaje usado.

Corresponden a las áreas de conocimiento de la lengua y conocimiento del tópico. Aquí se consideran dos características:

- i) características del lenguaje que pueden ser organizacionales (vocabulario, morfología, sintaxis, fonología y grafología) y
- ii) textuales (cohesión, organización retórica o conversacional); y características del tópico que se refieren al tipo de información del estímulo como personal, académica, técnica o de otro tipo.

**Características de la respuesta esperada**. En una situación de uso del lenguaje los participantes tienen expectativas acerca de las respuestas de los otros en relación con su discurso. Por otra parte en una prueba del manejo de un idioma la respuesta esperada consiste en el uso del lenguaje o en la respuesta que se está tratando de producir a través de las instrucciones, las tareas y el estímulo que se presenta. En este modelo se distingue la respuesta del sustentante de la respuesta esperada porque en ocasiones los participantes no responden de la manera prevista, sea porque no entienden las instrucciones, o porque deciden responder de otra manera. La respuesta esperada de la prueba tiene características correspondientes a las del estímulo y por lo tanto se distinguen:

- a) el formato en el que se produce, el cual puede ser descrito en términos del canal, forma, tamaño, tipo y grado de velocidad;
- b) tipo de respuesta que puede ser seleccionada como en las pruebas de opción múltiple, o producida siendo ésta limitada (una o pocas palabras) o extensa (más de dos palabras, una oración, o todo un discurso);
- c) grado de velocidad de la respuesta que es el tiempo que le toma al sustentante planear y efectuar su respuesta. Cuando el puntaje de la respuesta depende de la velocidad con la que responde se habla de una respuesta cronometrada; y
- d) lenguaje de la respuesta esperada, que se refiere a las mismas características que se mencionaron para el estímulo, es decir
  a. i) características del lenguaje que pueden ser organizacionales (vocabulario, morfología, sintaxis, fonología y grafología) y textuales (cohesión, organización retórica o conversacional); y

b. ii) características del tópico que se refieren al tipo de información que ofrece el estímulo como personal, académica, técnica o de otro tipo.

**Relación entre el estímulo y la respuesta**. Hace falta describir las características de la relación entre el estímulo y la respuesta porque en este aspecto hay tres factores que se pueden identificar y distinguir y por lo tanto pueden servir como criterios para evaluar la viabilidad o la validez de un examen. Las características son:

- a) reactividad que es el grado en el que el estímulo o la respuesta afectan directamente a los estímulos o las respuestas subsiguientes. En este sentido pueden ser de tres tipos:
  - i) tareas recíprocas, que son aquellas en las que el sustentante o el usuario del lenguaje se involucran en el uso de la lengua con su interlocutor y reciben realimentación acerca de qué tan correcta y relevante es su respuesta; la respuesta a su vez afecta al estímulo subsecuente producido por su interlocutor. La realimentación puede ser explícita o implícita por parte del interlocutor, distinguiendo la sola presencia de realimentación de la interacción entre los dos interlocutores (como ocurre en una conversación cara a cara);
  - ii) tareas no recíprocas son las que no implican realimentación, ni interacción entre los usuarios del lenguaje. Por ejemplo la lectura es una tarea del uso de la lengua en la que la respuesta del individuo no afecta al material leído o al interlocutor, que es el autor del texto. Otros ejemplos son tomar dictado o escribir una composición;
  - iii) tareas adaptativas que constituyen uno de los más recientes avances en el diseño y construcción de instrumentos de evaluación educativa. En este tipo de exámenes la respuesta a la tarea presentada ejerce influencia sobre la tarea que se ofrece subsecuentemente. La primera tarea típicamente es de dificultad media y si la respuesta a ésta es correcta, la siguiente tarea presentada será ligeramente más difícil. Por lo contrario si la respuesta es incorrecta la siguiente tarea será ligeramente más fácil y así sucesivamente con todas las tareas del

examen. Este es un ejemplo de interactividad, pero no necesariamente de la presencia de realimentación porque el sustentante puede saber o no que la dificultad de las tareas se modifica en función de sus respuestas. En un examen del manejo de idiomas en una entrevista por ejemplo el interlocutor puede disminuir la velocidad, la dificultad del vocabulario o hacer pausas para facilitar la comprensión auditiva de su interlocutor en el caso de percibir que éste tiene dificultades para comprender;

- b) el espectro de la relación entre estímulo y respuesta se refiere a la cantidad de información que el sustentante debe de procesar en el estímulo para dar una respuesta. El espectro puede ser:
    - i) amplio cuando tiene que procesar una gran cantidad de información como cuando tiene que responder a una tarea en la que se le pide que identifique la idea principal de un texto o de una conversación;
    - ii) o reducido cuando el sustentante tiene que procesar poca información para dar su respuesta como ocurre en gran parte de reactivos en las pruebas de opción múltiple de respuesta seleccionada;
- c) precisión de la relación se refiere al grado en el que la respuesta esperada se basa en la información ofrecida por el estímulo, o tiene que tomar elementos del contexto o incluso, en el conocimiento del tópico. Esta característica tiene dos aspectos:
    - ) directa cuando la respuesta esperada depende solamente de la información que ofrece el estímulo, como cuando se lee un fragmento de texto y se hace una pregunta acerca de la información proporcionada;
    - ii) indirecta cuando la respuesta esperada depende de información no proporcionada por el estímulo, como en el caso de una conversación en ambiente natural en el que la respuesta del interlocutor puede brindar información nueva, no exclusivamente relacionada con la pregunta.

La utilidad práctica que representa la clasificación de las características de las tareas del uso de la lengua queda manifiesta a partir de esta taxonomía de tareas que proporciona los criterios a considerar cuando se requiere construir o evaluar una prueba que mide el manejo de un idioma. Hasta aquí la caracterización de las tareas del uso del lenguaje de acuerdo con el modelo de Bachman y Palmer (1996). Enseguida se ofrece la clasificación de las variables que se relacionan con la persona que contesta la prueba.

#### 4.5.3.3. Las características de los sustentantes y la habilidad en el lenguaje.

En el amplio modelo de Bachman y Palmer se ofrecen los elementos para caracterizar al individuo y su uso del lenguaje. Ya sea que se trate de un usuario del lenguaje en su ambiente natural, o un sustentante de una prueba se están abordando las características personales, el conocimiento del tópico, el esquema afectivo y la habilidad en el lenguaje.

Estos autores conceptualizan la habilidad de una manera específica dependiendo del sujeto hablante y de la situación en la que se expresa y lo denominan uso del lenguaje el cual

> [...] puede ser definido como una creación o interpretación de significados intencionales en el discurso de un individuo, o como la negociación dinámica e interactiva de significados intencionales entre dos o más individuos, en una situación particular" (Bachman y Palmer, 1996, p. 61).

De esta manera la interacción del individuo y el ambiente implica un estímulo que evoca una respuesta, siendo el estímulo una situación determinada en la que el sujeto responde de manera congruente con sus propias características, pero en relación con las características del ambiente y lleva el análisis de la interacción a un plano más complejo que el de una respuesta condicionada, desde el modelo conductista, sin necesidad de postular estructuras gramaticales innatas, como en la propuesta de Chomsky (1971, citado en Santos Caicedo, 2007).

En la Figura 4.1 se presentan esquematizadas las interacciones de los componentes del uso del lenguaje del individuo y la ejecución de éste en la prueba. Las características del uso de la lengua, de las tareas de la prueba y de su administración representadas en el óvalo exterior

entran en contacto con la habilidad del sustentante en el idioma, representada por el conocimiento del tópico y el conocimiento del lenguaje y con las características personales a través del vínculo cognoscitivo constituido por las competencias estratégicas, siempre matizadas por el mundo afectivo del individuo.

El modelo de Bachman y Palmer no pretende ser una teoría lingüística, sino un marco conceptual para construir y usar exámenes del manejo de lenguas que permita también operacionalizar los elementos que deben ser considerados en la definición del constructo de una prueba. Es decir dependiendo del tipo de prueba que se quiere construir o evaluar el conocimiento de la lengua (que es lo que generalmente se desea medir), el conocimiento del tópico (que en ocasiones se trata de evitar para no sesgar los reactivos) y las características personales deben considerarse en varios momentos de la construcción y uso de la prueba para aumentar el grado de validez en general.

En los siguientes párrafos se explican los elementos con los que el modelo de Bachman y Palmer proponen caracterizar a los sustentantes y su habilidad en el lenguaje.

**Características de los individuos**: se refiere a las características que deben ser consideradas en términos de su contribución potencial a la utilidad de la prueba:

I) características personales: a) edad, b) sexo, c) nacionalidad, d) estatus de residente, e) lengua nativa, f) nivel y tipo general de educación, g) tipo y cantidad de preparación o experiencia previa con la prueba

II) II) conocimiento del tópico es el conocimiento que el individuo tiene de los aspectos culturales de la lengua tal y como se usa en el mundo real;

III) III) esquema afectivo es el correlato emocional o afectivo del conocimiento del tópico, es decir lo que el individuo siente con respecto al contexto cultural de la lengua, que puede facilitar o dificultar su respuesta.

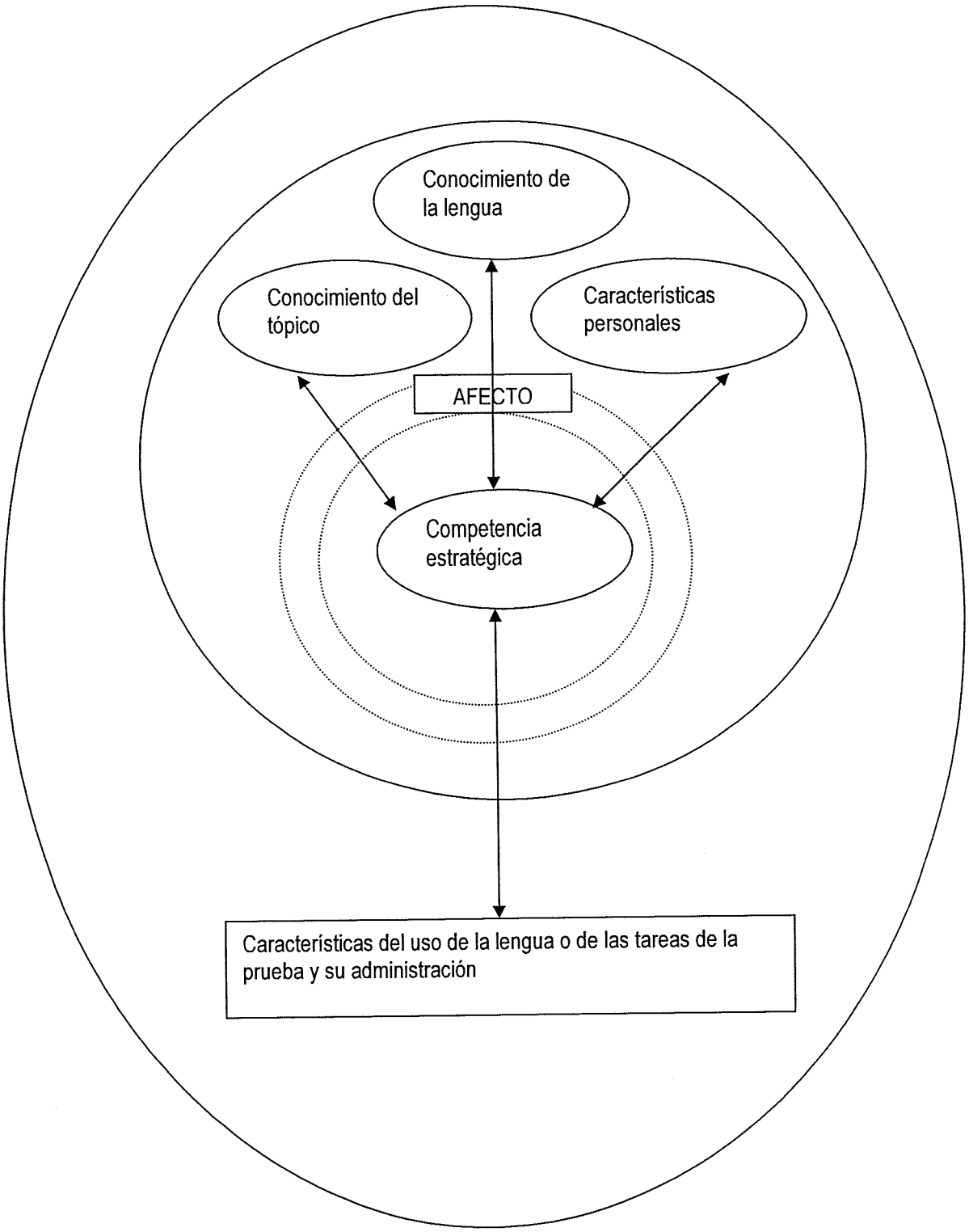



Figura 4.1. Algunos componentes del uso del lenguaje y ejecución en la prueba. Tomado de Bachman y Palmer, 1996, p. 63.

Habilidad en el idioma: a diferencia de las cuatro habilidades tradicionalmente hipotetizadas para explicar el manejo de una lengua, en este modelo la habilidad se define de manera precisa para permitir diferenciar la habilidad, de otras características de la persona que no estén relacionadas con la habilidad subyacente y que pudieran afectar los resultados de la prueba. Entonces la manera en la que se defina la habilidad en el idioma para una situación particular de prueba constituye el fundamento para la clase de inferencias que se hagan sobre los resultados que obtiene el sustentante. Como puede comprenderse al definir las habilidades de esta manera, se está definiendo el constructo que mide la prueba.

En el modelo de habilidad de Bachman (1990) la combinación de dos componentes proporciona a los individuos la habilidad o la capacidad para crear, o interpretar el discurso sea para responder a las tareas de una prueba o para responder en situaciones distintas de la prueba. A continuación se explican los dos componentes y los elementos que los integran:

Conocimiento de la lengua es la competencia en el idioma, o como dicen los autores del modelo "es el dominio de información en la memoria que está disponible para ser usado por las estrategias cognoscitivas del individuo, para crear e interpretar el discurso en el uso de la lengua (p. 67) y El conocimiento de la lengua puede a su vez ser de dos tipos:

- conocimiento organizacional que controla la estructura formal de la lengua para comprender o producir lenguaje el cual a su vez puede ser de dos tipos:
    - A) conocimiento de gramática que está implícito en la producción o comprensión de oraciones o de discursos y consta de:
        - i) conocimiento del vocabulario
        - ii) conocimiento de la sintaxis
        - iii) conocimiento de la fonología/grafología;
    - B) el conocimiento textual que es el conocimiento relacionado con producir o comprender textos sean hablados o escritos y que puede ser de dos tipos:

- i) conocimiento de cohesión que interviene en la producción o comprensión de las relaciones explícitamente marcadas entre las oraciones; y
- ii) conocimiento de organización retórica o conversacional que es el utilizado para producir o comprender el desarrollo organizacional de textos escritos o conversaciones.

b) conocimiento pragmático que permite crear o interpretar discursos a través de interpretar lo que se dice oralmente o por escrito de acuerdo con los significados, o las intenciones del interlocutor y con las características relevantes del uso del lenguaje. Hay dos áreas de conocimiento pragmático: conocimiento funcional y conocimiento sociolingüístico.

El conocimiento funcional incluye cuatro categorías de funciones: i) conocimiento de funciones ideacionales que le permiten al individuo expresar o interpretar significados en términos de su experiencia con el mundo real. Estas funciones incluyen el uso del lenguaje para expresar o intercambiar información acerca de ideas, conocimiento o sentimientos. Como ejemplos de este tipo de conocimiento están las descripciones, clasificaciones, explicaciones, expresiones de condolencia o de enojo; ii) conocimiento de funciones manipulativas que son las que facilitan el uso del lenguaje para producir un efecto en el entorno y son de tres tipos: A) funciones instrumentales que se realizan para hacer que otras personas hagan algo como solicitudes, peticiones, sugerencias, órdenes o advertencias B) funciones regulatorias que se usan para controlar lo que otras personas hacen como la formulación de reglas, reglamentos o leyes; C) funciones interpersonales que se usan para establecer, mantener y cambiar relaciones interpersonales, como decir un piropo, un cumplido, una disculpa, un insulto; iii) heurísticas que permiten el uso del lenguaje para ampliar el conocimiento acerca del entorno como el lenguaje que se usa para enseñar o para aprender, para resolver problemas, o para retener información; iv) conocimiento de funciones imaginativas que es el que subyace a la creación de un mundo imaginario o para extender el entorno con fines humorísticos o estéticos; como ejemplos pueden citarse los chistes o el uso de lenguaje figurativo y la poesía.

El conocimiento sociolingüístico es el que permite crear o interpretar el lenguaje que es apropiado para un uso de lenguaje en un contexto determinado. Esto incluye i) el conocimiento

de las convenciones que determinan el correcto uso de los dialectos, las variedades de lenguas naturales; ii) el conocimiento del registro que implica la discriminación de la manera apropiada para dirigirse al interlocutor, como por ejemplo a una autoridad, o en una conferencia, etc.; iii) las expresiones idiomáticas que no tienen un significado literal, sino cultural y se refieren a situaciones de uso general como por ejemplo "le salió el tiro por la culata"; iv) conocimiento de referencias culturales que expresan significados comunes a una colectividad, como los dichos populares.

II). la competencia estratégica que son un conjunto de estrategias metacognocitivas. La competencia estratégica consta de un conjunto de procesos ejecutivos de alto orden que proporcionan una función de manejo cognoscitivo en el uso del lenguaje y en otras actividades. Usar el lenguaje implica el conocimiento del tópico que tiene el individuo en interacción de su esquema afectivo y todas las áreas de conocimiento del lenguaje que mencionamos anteriormente. Desde el punto de vista de la evaluación del manejo de la lengua la conceptualizacion de una competencia estratégica, definida en términos de componentes cognoscitivos proporciona el fundamento necesario para diseñar y construir tareas de prueba potencialmente interactivas, así como para evaluar la interactividad de las pruebas. Hay tres áreas en las que operan estos componentes metacognoscitivos:

a) Establecer metas tiene que ver con la decisión que se toma acerca de lo que uno va a hacer e implica: A) identificar las tareas del uso de la lengua, o las tareas de la prueba; B) escoger una o más tareas disponibles y C) decidir si se completarán las tareas de la prueba. Ya que el propósito de una prueba del manejo de un idioma es provocar una situación típica específica del uso de la lengua pero se hace habitualmente en formas mucho más restringidas que las que encontraría el sustentante en el ambiente natural, conviene operacionalizar los componentes cognoscitivos para poder medirlos.

b) Ponderar o evaluar lo que se necesita para cumplir con la tarea y evaluar qué tan apto se siente el sustentante para realizarla. El sustentante debe: A) evaluar las características de las tareas para determinar si puede o quiere involucrarse en contestarlas y B) evaluar el propio conocimiento del tópico y del lenguaje para determinar si se tiene el conocimiento necesario para completar exitosamente la tarea; C) evaluar qué tan correcta o apropiada es la respuesta del

propio sustentante con respecto a la tarea de acuerdo con los criterios de corrección o propiedad percibidos.

c) Planeación es la estrategia que permite decidir cómo contestar la tarea. Implica tres aspectos: A) utilizar el conocimiento del lenguaje y el conocimiento del tópico para completar las tareas de la prueba exitosamente. B) formular uno o mas planes para instrumentar los elementos del conocimiento del lenguaje y del tópico para responder a la tarea; C) seleccionar un plan para empezar a responder las tareas de la prueba.

#### 4.5.3.4. La utilidad de las pruebas.

La visión de Bachman y Palmer favorece la concepción integral del fenómeno del uso de la lengua y su evaluación en un solo esquema, en el que la utilidad de una prueba es la columna vertebral de todo el proceso de planeación, construcción, uso e interpretaciones de las pruebas. Dos principios fundamentan esta aproximación: 1) la necesidad de una correspondencia entre la ejecución del sustentante en la prueba y el uso que se le da a la lengua en situaciones fuera de la prueba y 2) la definición clara y explícita de las cualidades de la prueba que son: confiabilidad, validez del constructo, autenticidad, impacto, interactividad y viabilidad. En los siguientes párrafos se explican los dos principios.

1. Correspondencia entre la ejecución del sustentante en la prueba y el uso que se le da a la lengua en situaciones fuera de la prueba. Es claro que para poder garantizar la validez de las inferencias acerca de la habilidad que una persona tiene en una lengua se requiere demostrar que existe una correspondencia entre el uso de la lengua que se da en la situación de prueba y el que se da en situaciones ajenas a la prueba. Para lograrlo primeramente se tiene que conceptualizar la situación de prueba como una muestra particular del uso de la lengua en el mundo real. A partir de esa conceptualización se pueden identificar y definir las características de las actividades y las tareas que el sustentante tiene que realizar en la prueba y evaluar si se corresponden con las características de las tareas que son propias del uso de la lengua. La Figura 4.2 ilustra esquemáticamente la correspondencia entre el uso del lenguaje en general y en la situación de prueba.

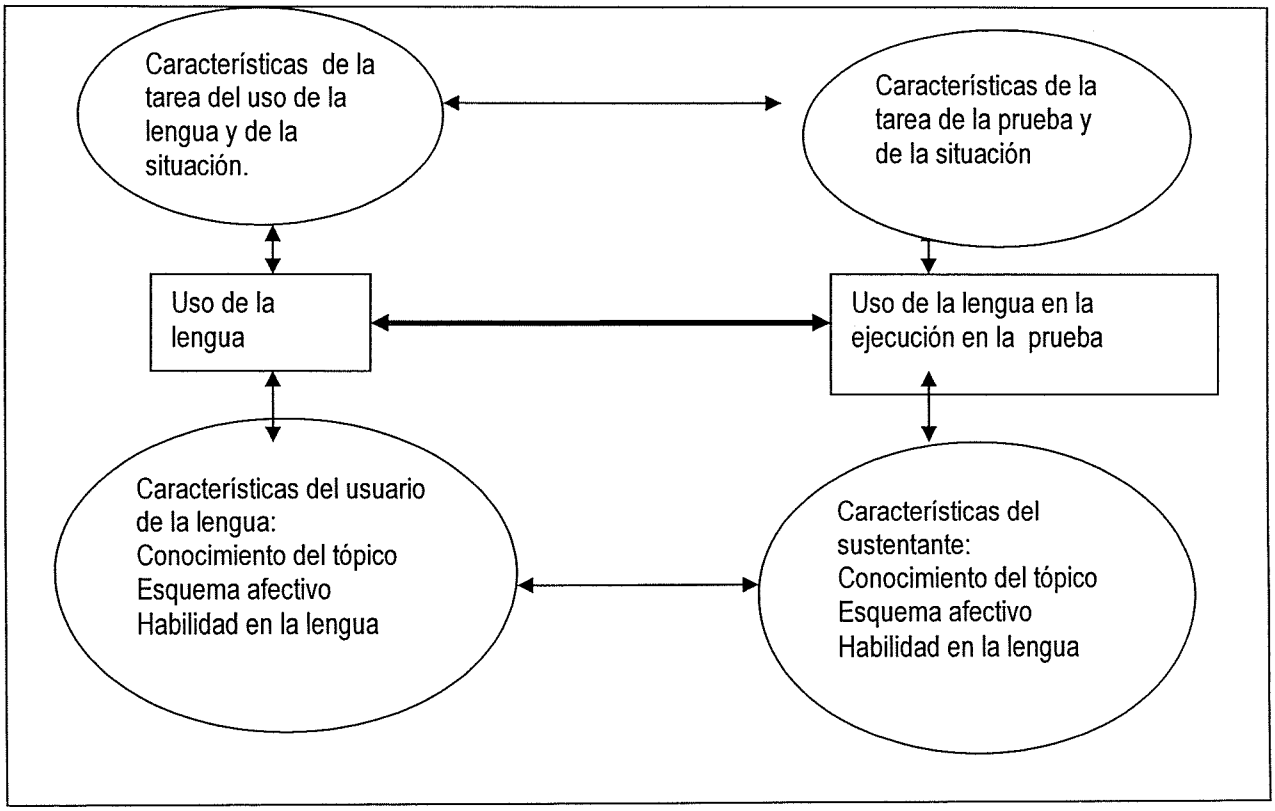



Figura 4.2. Correspondencia esquemática entre el uso de la lengua en la situación de prueba y otras situaciones, distintas de la prueba. Tomado de Bachman y Palmer, 1996, p. 12).

En el esquema de la Figura 4.2 se puede ver que la flecha que une los dos rectángulos centrales señala que los dos dominios: el uso de la lengua dentro y fuera de la situación de la prueba son correspondientes mutuamente. Para evaluar la validez de una prueba que mide el manejo de una lengua se tiene que demostrar que las características de la situación y las tareas de la prueba son esencialmente similares a las características de la situación y uso de la lengua, como lo ilustra la flecha entre los dos óvalos en la parte superior de la figura. En cuanto a las características de los individuos los óvalos de la parte inferior de la figura están unidos por una flecha de dos puntas para señalar la correspondencia que debería existir entre las características del usuario de la lengua en el mundo real y las del sustentante que contesta la prueba. Esta demostración es crítica ya que constituye el fundamento de las inferencias que se harán con respecto a los puntajes obtenidos en la prueba.

2. Cualidades de la prueba. Para que una prueba sea útil es decir que como instrumento de medición realmente mida aquello para lo que fue construido debe demostrarse que es confiable, que tiene validez de constructo, que las preguntas que la constituyen son relevantes y pertinentes con respecto a un dominio auténtico y que pueda dar sustento, tanto a las interpretaciones que se hacen de los puntajes obtenidos, como a las decisiones que se toman alrededor de éstos. Estas son las cualidades de las pruebas que han sido mencionadas por los expertos en evaluación de habilidades y conocimientos (Anastasi, 1977; Hambleton, 1988; Hughes, 1989; Popham, 1990; Messick, 1993; Alderson, 1993; Nitko, 1994, Bachman y Palmer, 1996 entre otros). En la literatura también se mencionan, aunque a veces de manera aislada y circunstancial (excepto en Messick, 1993, quien da cuenta de todas las cualidades de una prueba) otras cualidades que se relacionan con las consecuencias que la prueba tiene sobre el sustentante y sobre los usuarios. Así mismo, las variables personales del sustentante y los riesgos que representa la administración de la prueba a menudo son tratados de manera general junto con las variables que producen sesgo.

Bachman y Palmer recogen las cualidades mencionadas en la literatura sobre el amplio tema de la medición de las capacidades y habilidades de los estudiantes a través de instrumentos de medición y los aplican de manera sistemática a su modelo de la evaluación del manejo de los idiomas. Estos autores expresan la noción de utilidad de una prueba en función de la suma de esas cualidades, como se ve en el Recuadro 4.1 dejando claro que la utilidad de una prueba es una función de sus cualidades.

UTILIDAD: confiabilidad + validez de constructo + autenticidad + interactividad + impacto + viabilidad

**Confiabilidad.** Se refiere a la consistencia de una medida. Una prueba confiable brindará resultados consistentes de una aplicación a otra ya sea que su propósito sea jerarquizar a los sustentantes de mayor a menor habilidad, o evidenciar quiénes están por encima o por

debajo de un nivel dado; o si dos o más personas van a registrar o calificar algo de acuerdo con un criterio.

**Validez de constructo**. Es la significación y la propiedad de las interpretaciones que se hacen a partir de los puntajes obtenidos en una prueba. Un constructo en el contexto de la medición del dominio de un idioma es la definición específica de una habilidad, la cual da origen y sentido al diseño de una prueba, o de una tarea de la prueba, con la finalidad de efectuar interpretaciones y generalizaciones sobre esa habilidad.

**Autenticidad**. Para que una prueba sea útil, tiene que ser auténtica. Es decir las tareas que el sustentante debe realizar en la prueba deben ser esencialmente iguales a las que encontraría en el ambiente natural, fuera de las condiciones de evaluación. Curiosamente esta cualidad se ha dado por entendida y se trata de cumplir en los exámenes que los maestros cotidianamente ponen a sus alumnos en las escuelas, pero la literatura especializada en el tema la aborda de manera poco sistemática. Generalmente se menciona cuando se señalan los sesgos que pueden tener los exámenes, o el riesgo de generalizar las interpretaciones sobre los resultados a poblaciones, o a dominios que quedan fuera del alcance de lo que las pruebas realmente miden. Pero en definitiva es una característica que debe tomarse con toda seriedad puesto que constituye la directriz de la definición de las tareas de la prueba y la amplitud de la generalización de los resultados. Tiene la función de proyectar al sustentante un panorama que confiere credibilidad a las tareas, lo que hace que se perciban como relevantes o no relevantes con respecto del dominio que está siendo medido. La autenticidad es la cualidad de la prueba que hace que el sustentante interactúe con ella de una manera adecuada.

**Interactividad**. De la autenticidad se desprende la interactividad de una prueba. Sea que se considere abiertamente o no en el diseño de la prueba, el sustentante establece una relación con el instrumento de medición y puede percibir si las tareas que está realizando son pertinentes, o no lo son. Desde luego los sustentantes comparten entre ellos algunas características y otras son totalmente distintas, pero la característica de interactividad de los reactivos actúa sobre el grado de involucramiento que los sustentantes tienen con la prueba. Cuando se planea la interactividad de los reactivos tienen que tenerse presentes las características culturales de los sustentantes, porque esta cualidad se relaciona con la

capacidad que tiene el reactivo de instigar la habilidad de los sustentantes en el uso del idioma, de manera que la competencia estratégica, el conocimiento contextual y el esquema afectivo de cada sustentante pueden determinar las respuestas diferenciales entre estudiantes de diferentes contextos culturales. Más adelante explicaremos cada uno de estos factores.

**Impacto**. El impacto de evaluar a los estudiantes repercute a dos niveles; a un nivel micro afecta al individuo y al contexto inmediato. A un nivel macro afecta al sistema educativo y a la sociedad. En cuanto al nivel micro en la literatura sobre evaluación de idiomas los autores tratan con frecuencia el tema denominado en inglés *washback* que es una de las formas en las que la evaluación tiene un impacto más allá del individuo que sustenta la prueba y significa que, cuando los exámenes están bien hechos ayudan a mejorar el aprendizaje de la lengua; por lo contrario cuando hay sesgos de cualquier tipo pueden obstaculizar el avance del proceso de la enseñanza y el aprendizaje (Heaton, 1988, Hughes, 1989). A nivel macro los resultados de las pruebas pueden tener implicaciones en las percepciones o actitudes de la sociedad acerca de un tema en particular. Por ejemplo en las pruebas nacionales los puntajes bajos obtenidos por los estudiantes pertenecientes a los sectores marginales de la población (como pueden ser los hijos de inmigrantes) han creado la percepción de que la raza es un factor de bajo rendimiento siendo que es más probable que los exámenes mismos estén midiendo de manera sesgada las habilidades en poblaciones que presentan profundas diferencias culturales. De manera similar los resultados de las pruebas educativas pueden originar políticas educativas que sean benéficas para unos sectores de la población, pero no para otros.

**Viabilidad**. La naturaleza de esta cualidad es distinta de la de las anteriores ya que aquellas se refieren a los usos que se dan a los puntajes obtenidos en las pruebas y las cualidades prácticas tienen que ver con las formas en las que va a ser aplicada, así como con las posibilidades de uso que tiene la prueba. La viabilidad no se refiere solamente a los procedimientos de su aplicación; también se refieren al grado en el que las demandas de las especificaciones de la prueba pueden cumplirse dentro de los límites de los recursos existentes. Los recursos pueden ser de tres tipos: humanos, materiales y de tiempo. Estos tres criterios de la adecuación práctica de una prueba pueden ser definidos operacionalmente y por lo tanto, son susceptibles de ser medidos.

# 5. Validación del contenido del EXEDII

En los capítulos anteriores de este trabajo de tesis se ha venido presentando la información que permite contextualizar y justificar la investigación de las evidencias de validez EXEDII. Se planteó la problemática de la que se deriva la estrategia general de la indagación y se ofrecieron los argumentos teóricos que la sustentan. También se presentó la información necesaria para comprender la naturaleza del instrumento que se está evaluando. Llegados a este punto los capítulos 5 y 6 presentan toda la información relativa a la estrategia metodológica de búsqueda de evidencias de validez. En este capítulo se presenta el proceso de acopio de evidencias de la validez de constructo y en el siguiente el de evidencias de validez de contenido.

## 5.1. Planteamiento del problema.

En el Capítulo 4 de este trabajo de tesis se presenta una revisión somera del tema de la validez de los instrumentos de medición psicológica y educativa, en donde se explica la forma en la que se han abordado los estudios de validación de pruebas. El método que más frecuentemente se sigue en los estudios de validación de contenido implica el juicio de un grupo de expertos que comparan los reactivos del instrumento, contra un criterio que generalmente es un currículo (Anastasi, 1977). El EXEDII es un examen criterial de certificación del inglés, por lo que la metodología seguida para su construcción implicó su alineación a un currículo (Nitko, 1994; Hambleton, 1988, Popham, 1990). El currículo (o criterio) en este caso es el tercer nivel del curso de inglés que estaba vigente en lo que era la Escuela de Idiomas de la UABC. Por ello, se pensó que volver a revisar si efectivamente los reactivos estaban diseñados de acuerdo con los objetivos del curso sería una actividad ociosa, puesto que esto se hizo durante la construcción del examen.

Por otra parte, la investigación documental de la temática sobre la validez de la medición del lenguaje puso en evidencia que por la naturaleza del objeto de estudio por un lado, y dado que la validez es un concepto unitario, la búsqueda de evidencias de validez del EXEDII debería de abordarse con un enfoque distinto al que prevalece en muchos estudios de

validación de pruebas, y en el que se superasen algunos problemas derivados no solamente de la metodología utilizada, sino de la concepción misma del lenguaje y su medición.

A propósito de lo anterior, en 1996 Bachman y Palmer propusieron un modelo para la construcción y uso de pruebas que miden el manejo de idiomas. Se trata de un modelo que parte de un enfoque comunicativo que propone estrategias para el manejo empírico de los conceptos teóricos fundamentales, integrando los resultados alrededor del concepto de la utilidad de la prueba, en lugar de ofrecer evidencias de validez de manera aislada.

En el apartado 4.2.3.4 se trata con mayor amplitud el concepto de la utilidad de las pruebas, que implica dos principios: 1) la necesidad de una correspondencia entre la ejecución del sustentante en la prueba y el uso de la lengua en situaciones fuera de la prueba y 2) la definición clara y explícita de las cualidades de la prueba, que son: confiabilidad, validez del constructo, autenticidad, interactividad, impacto y viabilidad.

El estudio de validación de contenido del EXEDII responde al primer principio, a través de la respuesta a tres de las preguntas de la investigación que reporta esta tesis. Las preguntas se retoman en este capítulo para ayudar a la comprensión de la estrategia seguida:

> ¿Cuáles son las características del dominio de uso de la lengua meta (DULM) que debería medir el EXEDII?
> ¿Cuáles son las características de las tareas del dominio de uso de la lengua meta (DULM) que debería medir el EXEDII?
> ¿Los reactivos del EXEDII constituyen una instancia del uso del inglés de los estudiantes que egresan de la UABC y si ese es el caso, de qué manera estos datos aportan evidencias de validez de contenido y de constructo de la prueba?

Así mismo, la estrategia responde a dos de las cualidades de la utilidad de la prueba que son la autenticidad y la interactividad de sus reactivos. En el siguiente apartado se explican los pasos seguidos para recabar evidencias de validez de contenido del EXEDII.

## 5.2 Elaboración del criterio de contenido.

La respuesta a las dos primeras preguntas de investigación planteadas para la validación del contenido del EXEDII requirió la instrumentación de un procedimiento realizado en dos fases: 1) la investigación de las características del DULM y de las tareas que lo componen para formular, con base en ellas el criterio de contenido, 2) la comparación de los reactivos del EXEDII contra ese criterio. La primera fase consta de tres actividades; el resultado de las dos primeras constituye un marco de referencia para acotar las tareas del dominio a medir, adecuándolas al contexto sociocultural de los sustentantes. Así, en la primera actividad se indagaron las características que debería tener el DULM, para lo cual se preguntó directamente a los profesores que conocen las características de los sustentantes y del contexto cultural en el que típicamente se desenvuelven. En la segunda actividad se investigaron las características de las tareas del DULM preguntando a los expertos en docencia del inglés como lengua extranjera, en el contexto de Baja California. La tercera actividad de la Fase 1 fue revisar los criterios de los estándares internacionales elaborados por expertos, pero en el contexto de otros países. La Fase 2 del estudio constó de una actividad que integra la información de las tres actividades anteriores en una conceptualización del DULM.

## 5.3. Metodología

La tarea de formular una conceptualización del dominio que mide un examen como el EXEDII parte de la necesidad de buscar los indicadores más relevantes del uso de la lengua meta, para integrarlos en una descripción operacional de las tareas que pertenecen al dominio del uso de la lengua a medir, que sean congruentes con las características de los sustentantes, en las situaciones particulares en que probablemente se darían (Bachman y Palmer, 1996).

Un ejemplo de dominio de uso de la lengua pueden ser las diferentes instancias de comunicación que enfrenta un turista. Pero un examen de certificación, como el EXEDII no puede evaluar solamente ese dominio, sino que debe incluir otros dominios igualmente amplios. No obstante, por razones del tamaño de la prueba, es necesario elegir una muestra representativa de las instancias de esos dominios, la cual constituye el DULM, mencionado párrafos arriba. Para tener una idea clara del tipo de tareas que constituyen un DULM es

necesario definir algunas características de la situación del uso de la lengua y del tipo de personas que enfrentarían esas tareas. Así, el DULM del EXEDII estaría conformado por una muestra de las instancias que enfrentaría un estudiante que egresa de la UABC, cuya lengua nativa es el español y que acude por ejemplo como turista, a una cafetería en un país de habla inglesa y desea ordenar el desayuno. Otros dominios de uso de la lengua podrían ser del ámbito escolar, laboral, o social. Es claro que la descripción de las características de las situaciones y de las personas debe ser más detallada que el DUL, pero sin llegar a un nivel de detalle que impida después efectuar generalizaciones de los resultados a otros sustentantes con características similares. El criterio que debe regular hasta qué grado se lleva la especificidad de las características depende del propósito del examen (Hambleton, 1988; Popham, 1990; Nitko, 1994; Bachman y Palmer, 1996).

**Criterio de los expertos.**

**Primera actividad: Grupo de Enfoque.**

**Propósito:** Recabar las opiniones de una muestra intencional de docentes universitarios acerca de dos tópicos: a) EXEDII, como instrumento de medición y b) los conocimientos y habilidades en inglés que los estudiantes de la UABC deben tener para egresar de la universidad y ejercer su carrera, o continuar estudiando.

Se recurrió al Grupo de Enfoque por constituir una técnica poco estructurada que favorece el intercambio de opiniones, sin necesidad de llegar a una conclusión porque lo que se pretende es favorecer respuestas espontáneas, que posiblemente no se obtendrían en un formato estructurado, o menos flexible (Loera Varela, 2000; Álvarez-Gayou, 2005).

**Participantes:** Se invitó a 16 docentes en activo a participar en el Grupo de Enfoque. Los 13 que aceptaron la invitación tenían las siguientes características:

Cuatro docentes de inglés en activo que habían dado al menos un curso a estudiantes de nivel intermedio. Siete docentes de escuelas o facultades de la UABC, con experiencia profesional en el área de su carrera y que además manejaban el inglés al menos al nivel intermedio. Un profesor representó a cada una las áreas académicas de las carreras de la UABC: económico-administrativas, químico-biológicas, salud, ingenierías, físico-matemáticas, ciencias sociales y humanidades. Una lingüista, un psicómetra y la autora de esta tesis, quien

fungió como Facilitadora en el Grupo de Enfoque organizando y centrando el tema durante la sesión. Además se ocupó de instrumentar las condiciones para realizar la video grabación de la sesión; así mismo, tomó notas a lo largo de la sesión y realizó el análisis de los datos obtenidos.

**Procedimiento:** Previo al Grupo de Enfoque, los participantes tuvieron la oportunidad de responder el EXEDII en las mismas condiciones en que lo hacen los estudiantes egresados de la UABC. Las preguntas que se formularon a los panelistas fueron: ¿qué debería saber de inglés un estudiante que egresa de la UABC? y ¿cuál es su opinión del EXEDII como instrumento para medir el manejo del inglés en los estudiantes de la UABC? La sesión tuvo una duración de tres horas y fue videograbada para conservar un registro de las opiniones e interacción de los participantes.

**Análisis de los datos del Grupo de Enfoque:** No se hizo una transcripción de la video grabación por considerarse innecesario ya que el propósito de la actividad no era el de analizar a profundidad el discurso. En lugar de ello se revisó la videograbación varias veces y se anotaron los temas tratados por los panelistas. Los temas anotados se organizaron en categorías conceptuales que agrupan varias ideas similares. Para complementar estas categorías se realizaron entrevistas a dos expertos en docencia del inglés en instituciones fuera de la UABC, buscando captar las opiniones de profesores que conocieran las características de los jóvenes profesionistas que desean aprender las competencias que demanda el mercado laboral de la región. En los siguientes párrafos se describe el procedimiento para efectuar las entrevistas y la forma de analizar los datos en ellas obtenidos, los cuales se ilustran en la Figura 5.1. Posteriormente en la Figura 5.2 se muestran las dos preguntas y el análisis de las respuestas obtenidas en las dos actividades que implican la interrogación a los expertos docentes, para ilustrar la integración de los resultados de esas actividades.

**Segunda actividad: entrevistas.**

**Fuentes de información.** Respuestas a las entrevistas a dos expertos en la docencia del inglés, fuera del ámbito de la UABC.

**Propósito**: Obtener respuestas detalladas sobre la opinión de los especialistas en la docencia del inglés, para contar con elementos para construir la conceptualización del DULM.

**Participantes:** dos docentes de inglés en activo, uno de ellos en la Normal Estatal y el otro, director de una academia en donde se imparte el inglés, de acuerdo con el programa "Inglés para propósitos específicos".

**Procedimiento**: se entrevistó por separado a los dos profesores de inglés. El formato de la entrevista fue semiestructurado y se formuló una de las preguntas del grupo de enfoque: ¿qué debería saber de inglés un estudiante que egresa de la UABC? No se formuló la otra pregunta que se hizo a los panelistas del Grupo de Enfoque porque uno de los entrevistados no tuvo oportunidad de conocer el EXEDII. La información recabada en las dos entrevistas fue analizada en dos partes, primeramente se escuchó la grabación un número de veces, para anotar las ideas generales expresadas. De manera similar a lo que se hizo con la información del Grupo de Enfoque, las ideas anotadas fueron agrupadas por similitud temática en categorías conceptuales. La Figura 5.1 ilustra el proceso y las categorías formuladas a partir de las respuestas de los entrevistados.

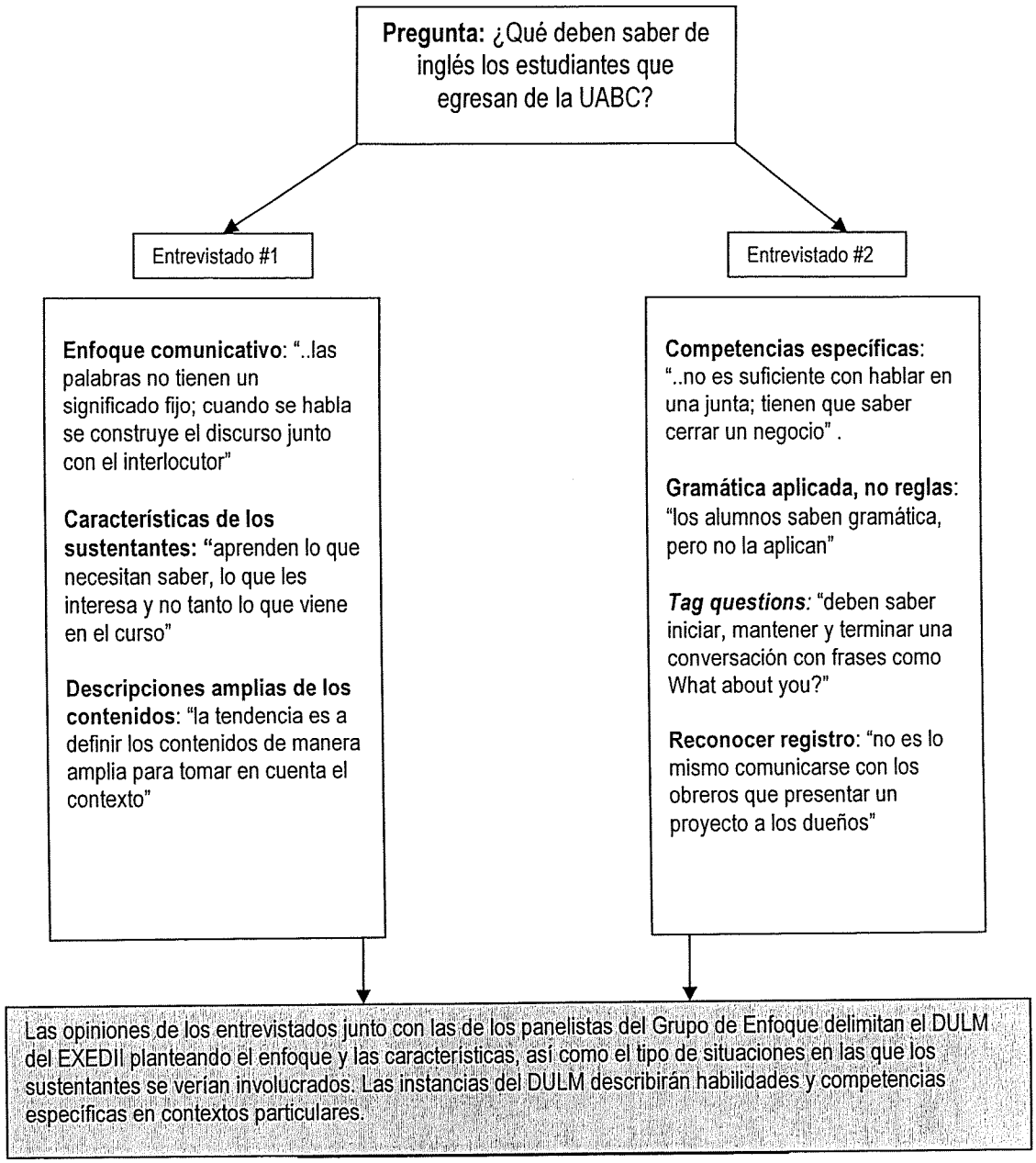



Figura 5.1. Esquema del análisis de información obtenida en entrevistas a dos expertos en docencia del inglés.

Como síntesis de las dos actividades anteriores se integraron las ideas expresadas por los expertos en tres grandes categorías que aparecen en el recuadro del extremo derecho de la Figura 5.2, en la que además se incluyen las preguntas de investigación a las que responden estas actividades.

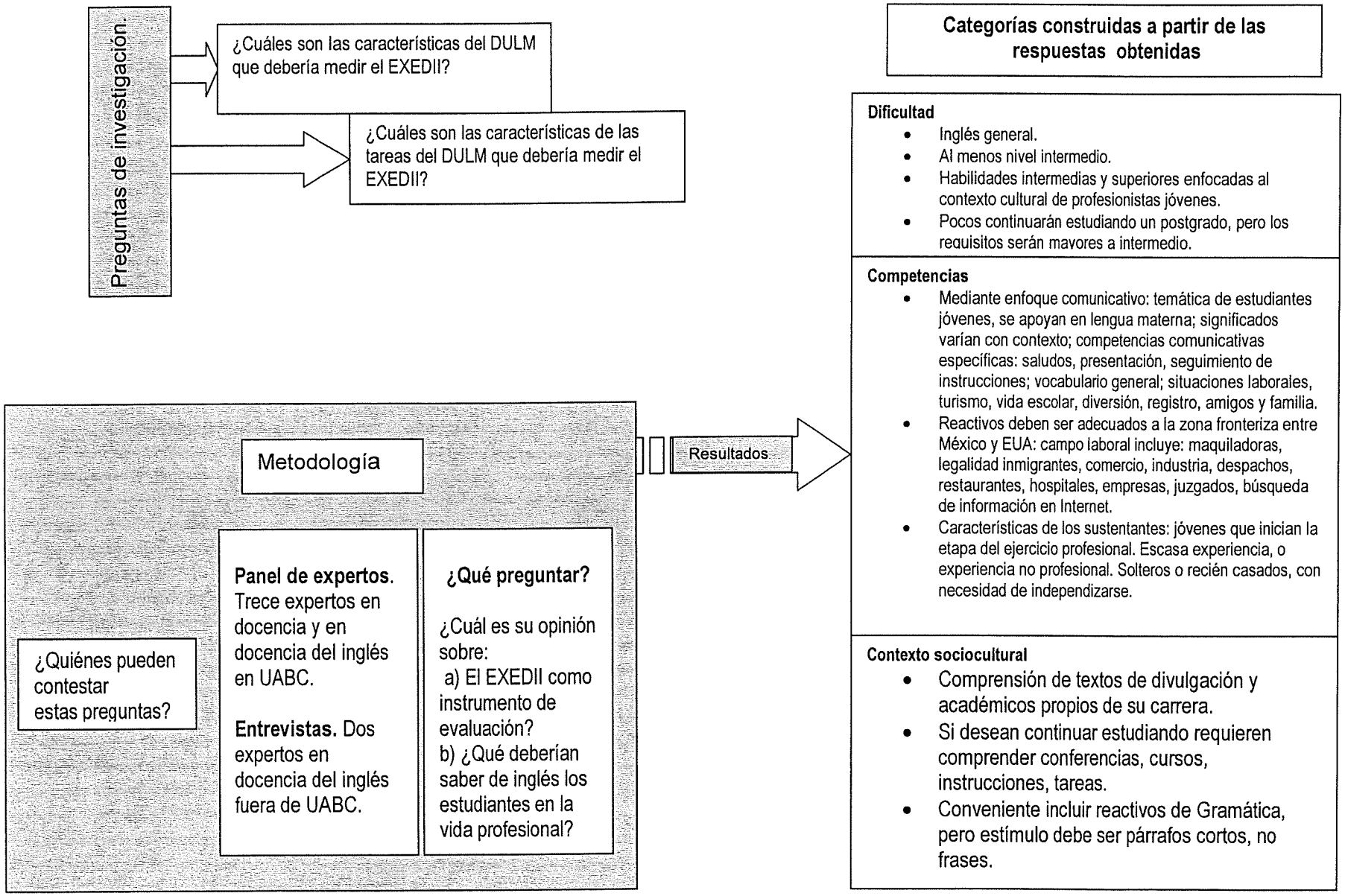



Figura 5.2 Esquema de las preguntas de investigación y método para contestarlas: resultados del Panel 1 y entrevistas. En el Recuadro derecha se muestran las categorías.

**Tercera actividad:** Revisión de cuatro estándares internacionales.

**Fuentes de información**. Se consultaron cuatro fuentes de información, todas ellas en el marco de los criterios internacionales de competencia al nivel intermedio del inglés como lengua extranjera: a) Consejo Americano de Profesores de Lenguas Extranjeras (*American Council of Teachers of Foreign Languages*, ACTFL, por sus siglas en inglés); b) Asociación de Profesores de Lenguas de Europa (*Association of Language Teachers of Europe, ALTE* por sus siglas en inglés); c) Universidad de Cambridge, Inglaterra y c) el programa de Certificación del Inglés como Lengua Extranjera, de la Universidad de Buenos Aires, Argentina (CILE).

Las razones por las que se eligieron estos organismos tienen que ver con el amplio uso que se hace de los estándares internacionales en los centros de enseñanza y evaluación de lenguas, tanto en América como en Europa y Asia. Así mismo, es notorio el impacto que tiene el sistema de exámenes de la Universidad de Cambridge (Inglaterra) sobre los criterios de aceptación de estudiantes en las universidades de los tres continentes mencionados. Finalmente, se eligió el CILE por ser una institución de enseñanza y certificación de idiomas, en el que el inglés se enseña como lengua extranjera, teniendo como lengua nativa al español, en un intento de homologar las circunstancias del EXEDII con las de una universidad latinoamericana.

**Propósito:** comparar los estándares del nivel intermedio de cuatro organizaciones internacionales para identificar las habilidades, competencias o conocimientos pertinentes al nivel que pretende medir el EXEDII.

Las organizaciones consultadas se identifican en lo sucesivo por sus siglas en inglés. De cada una de ellas se incluyeron en este análisis los niveles que cubren el nivel intermedio, quedando como se explica a continuación: a) ACTFL) en sus niveles intermedio bajo e intermedio medio; b) ALTE en sus niveles A1 y B1, en los rubros de dominio del inglés con habilidades generales, habilidades típicas de estudiantes, habilidades para el trabajo y para contextos turístico-sociales; c) CILE, en cuanto a las competencias mínimas de aprobación del programa de Certificación del Inglés como Lengua Extranjera, en sus descripciones de las etapas pre-intermedia e intermedia; d) la Universidad de Cambridge de acuerdo con los criterios de aprobación de los exámenes *Preliminary English Test* (*PET*) y *First Certificate in English (FCE*), ambos considerados como manejo intermedio bajo, e intermedio del inglés

como lengua extranjera, de acuerdo con los criterios de los exámenes ESOL *(English for Speakers of Other Languages)* de esa universidad (ALTE, 2007).

**Procedimiento**: se buscaron en Internet los estándares americanos (ACTFL) y los europeos (ALTE), así como los de un programa latinoamericano para certificar el dominio del inglés como lengua extranjera (CILE) y finalmente, se buscaron los requisitos de acreditación de dos exámenes de la Universidad de Cambridge en el nivel intermedio (PET y FCE). Con la información obtenida se construyó una tabla de dos entradas que se muestra en el siguiente apartado.

**5.3. Resultados**: la información recabada en los estándares de las organizaciones consultadas se organizó en las tablas 5.1 a 5.8 para facilitar la comparación de los contenidos. En las tablas se muestran fragmentos de las descripciones de las competencias correspondientes al nivel intermedio de los cuatro organismos consultados. Por razones de espacio las descripciones no se muestran completas, además de que algunas de las competencias descritas son irrelevantes al EXEDII, dadas las características de sus sustentantes. En las tablas mencionadas se incluyen las frases que expresan información de acuerdo con dos criterios: a) tareas que sobrepasan el nivel de sobrevivencia (inferior a intermedio) y que son afines a las características de los sustentantes del EXEDII y b) tareas que implican habilidades receptivas (lectura y comprensión oral, de acuerdo con la clasificación tradicional), así como habilidades productivas del mismo nivel.

La anterior selección se hizo con base en las características del EXEDII que es un examen que valora el nivel intermedio en sus tres subniveles (bajo, medio y alto), pero no mide las habilidades productivas (discurso hablado y escritura). Las frases que aparecen en cursivas fueron incorporadas a la conceptualización del DULM, que está redactado en términos de competencias, a pesar de que los nodos de los reactivos del EXEDII están redactados en términos de habilidades. Consideramos que esto es plausible porque aunque los conceptos de habilidad y de competencia tienen distinto alcance, siendo el más específico el de habilidad, de acuerdo con Elliot (1993, citado en Moreno Bayardo, 1998) el desarrollo de una habilidad determinada está articulada con otras habilidades. Por ello, aunque los objetivos educativos, o los procesos de evaluación se enfoquen en una habilidad particular, el desglose no implica que se conciba la habilidad en un vacío, sino dentro de una estructura de habilidades. A continuación se presentan las tablas mencionadas arriba y al final de ellas, la conceptualización del DULM, construida con los contenidos que aparecen en itálicas en las tablas.

Tabla 5.1. Descripción de algunas de las competencias/habilidades mínimas requeridas para el nivel intermedio del manejo del inglés de acuerdo con ACTFL, ALTE, el CILE de la Universidad de Buenos Aires y los exámenes ESOL de la Universidad de Cambridge.

| Organismo consultado | Competencias/habilidades típicas en contexto |
|---|---|
| | Habilidades generales |
| ACTFL<br>Intermedio bajo | El alumno es capaz de desarrollar con éxito un número limitado de tareas comunicativas no complicadas, *en situaciones sociales cara a cara*...intercambios concretos en tópicos predecibles, necesarios para la sobrevivencia en la cultura meta... *sobre información personal básica, familia, actividades diarias*. Conversación reactiva. |
| ACTFL<br>Intermedio medio | *Habilidad para manejar con éxito una variedad de tareas comunicativas no complejas en situaciones sociales cara a cara*...conversación... limitada a intercambios predecibles sobrevivencia en la cultura meta... expresa información de su persona... Funciona de manera reactiva respondiendo a preguntas directas. Si se requiere realizar funciones del nivel avanzado... *dificultad al hilar ideas, manipular tiempos* y aspectos, circunlocución (uso innecesario de palabras para expresar una idea)...se expresa... combinando y recombinando elementos de lenguaje conocidos y utilizando los estímulos conversacionales del interlocutor para elaborar frases cortos y algunas oraciones largas...en la búsqueda del vocabulario adecuado y las formas apropiadas para expresarse. Debido a errores de precisión, falta de vocabulario, malas interpretaciones, pero se dan a entender, sobre todo si el interlocutor es paciente y tiene experiencia con hablantes no nativos. |
| ALTE (2006)<br>Nivel 1 (A2) | Lectura: *información directa en área conocida como productos, señales...*<br>Auditiva/oral: expresa *[comprende] opiniones simples o requisitos en un contexto conocido...* |
| ALTE (2006)<br>Nivel 2 (B1) | Lectura: *información de rutina y artículos; significado general de información...*<br>Auditiva/oral: expresa *[comprende] opiniones en asuntos abstractos/culturales de manera limitada...* |
| CILE II<br>Etapa<br>preintermedia | *El estudiante adquiere mejor comprensión de situaciones y textos del primer nivel y logra manejar una más variada gama de situaciones cotidianas y de turismo; extrae información básica, entiende actitudes, opiniones y deseos.*<br>*En cuestiones académicas puede comprender textos de la especialidad.*<br>*En el campo laboral puede [comprender] el intercambio de opiniones [entre] colegas en temas conocidos y [comprender la orientación] a clientes en el ámbito de su conocimiento.* |
| CILE III<br>Etapa<br>Intermedia | *Conoce estructuras gramaticales principales, las utiliza con cierta confianza y precisión adecuada a la situación*. Soltura en el empleo de vocabulario variado.<br>*Tiene conciencia de registro formal e informal y comprende a hablantes nativos de una variedad conocida. Entiende diferentes puntos de vista en una conversación* y puede interactuar expresando hipótesis... formular preguntas y realizar presentaciones breves, *tomar apuntes en clases... de textos escritos... conocimiento básico del tema.*<br>En el ámbito laboral puede realizar tareas de oficina, intercambiar opiniones... |

Nota: Las habilidades/competencias seleccionadas para el DULM aparecen en itálicas.

Tabla 5.2. Descripción de algunas de las competencias mínimas requeridas para el nivel intermedio del manejo del inglés de acuerdo con ALTE: Habilidades de estudiantes y para el trabajo.

| | |
|---|---|
| Competencias/habilidades típicas de estudiantes | |
| ALTE (2006)<br>Nivel 1 (A2) | Lectura: significado general de un libro de texto simple leyendo muy despacio.<br>Auditiva/oral: [*comprende] opiniones simples usando expresiones como "I don't agree".* |
| ALTE (2006)<br>Nivel 2 (B1) | Lectura: *instrucciones, mensajes básicos catálogos computarizados de biblioteca.*<br>Auditiva/oral: *entiende instrucciones en clases y las tareas que deja el profesor.* |
| Competencias/habilidades típicas en el trabajo | |
| ALTE (2006)<br>Nivel 1 (A2) | Lectura: *[comprende] reportes…tema predecible en su área de conocimiento, si tiene tiempo suficiente.*<br>Auditiva/oral: expresa y *comprende requerimientos simples en [su] área de trabajo.* |
| ALTE (2006)<br>Nivel 2 (B1) | Lectura: *entiende significado general de cartas no rutinarias y artículos teóricos dentro de su área de trabajo.*<br>Auditiva/oral: ofrece [*comprende] ayuda/consejo a clientes en asuntos simples dentro de su área de trabajo.* |

Nota: Las competencias/habilidades seleccionadas para el DULM aparecen en itálicas.

Tabla 5.3. Descripción de algunas de las competencias/habilidades mínimas para el nivel intermedio del manejo del inglés de acuerdo con ALTE y los exámenes *Preliminary English Test (PET)* y *First Certificate in English (FCE)* de la Universidad de Cambridge.

| | |
|---|---|
| Competencias/habilidades típicas sociales y turísticas | |
| ALTE (2006) Nivel 1 (A2) | Lectura: *[comprende] información directa: etiquetas de alimentos, menús, señales en caminos, mensajes en máquinas para cambiar dinero.*<br>Auditiva/oral: expresa *[comprende] gustos/disgustos en contextos conocidos; me gusta/no me gusta.* |
| ALTE (2006) Nivel 2 (B1) | Lectura: *comprende artículos sencillos en periódicos, cartas rutinarias en hoteles, opiniones personales.*<br>Auditiva: *comprende limitadamente opiniones en temas abstractos/culturales y comprende detalles de opinión o inconvenientes.* |
| Competencias/habilidades de Sobrevivencia | |
| PET | Lectura: *anuncios, señales, noticias, textos cortos sobre hechos concretos; escanear textos para información específica.*<br>Auditiva: *conversaciones sencillas, anuncios, [idea general de] noticias en la radio…* |
| | Competencias/habilidades para el comercio, industria, instituciones educativas |
| FCE | Lectura: *comprende textos informativos y de interés general; comprensión del tema general…*<br>Auditiva: *comprende conversaciones, anuncios…* |

Nota: Las habilidades/competencias seleccionadas para el DULM aparecen en itálicas.

Dado que la descripción de los estándares consultados no está organizada en términos de cuatro habilidades (dos productivas y dos receptivas) las descripciones de la Tabla 5.1 se integraron en las tablas 5.4, 5.5, 5.6 y 5.7, reagrupándolas en las áreas de medición del EXEDII: comprensión auditiva, lectura y gramática, cuando los estándares incluyen esta última.

Tabla 5.4 Integración de la información de la tabla 5.1 de acuerdo con las áreas de medición del EXEDII: comprensión auditiva y lectura según ACTFL.

| Organismo consultado | Criterio lectura | Criterio auditiva |
|---|---|---|
| ACTFL Nivel intermedio | • Lee consistentemente comprendiendo textos simples interconectados acerca de una variedad de necesidades básicas y sociales.<br>• Tales textos son lingüísticamente sencillos y tienen una clara estructura interna subyacente. Proporcionan información básica de la que el lector requiere hacer una mínima cantidad de suposiciones y son de interés personal o conocimiento del lector.<br>• Ejemplos de ello son descripciones cortas y concretas de personas, lugares y textos de divulgación general. | • Capaz de entender discurso conformado por frases cortas que consisten en recombinaciones de frases aprendidas en una variedad de tópicos.<br>• El contenido se refiere a contextos personales, necesidades personales y sociales, convenciones y algunas tareas un poco más complejas como alojamiento, transporte y compras.<br>• Otras áreas de contenido incluyen algunos intereses personales, actividades y una diversidad de instrucciones y direcciones.<br>• Las situaciones de uso se refieren a conversaciones cara a cara, aunque son capaces de [comprender] conversaciones cortas y rutinarias por teléfono, y algunos discursos deliberados como anuncios y reportes en medios de comunicación.<br>• La comprensión tiende a ser incompleta. |

Tabla 5.5 Integración de la información de la Tabla 5.1 de acuerdo con las áreas de medición del EXEDII: Comprensión auditiva y Lectura, según ALTE.

| Organismo | Criterio lectura | Criterio auditiva |
|---|---|---|
| ALTE (2006) Nivel 2 (B1) | • General: entiende información de rutina y artículos; significado general de información no rutinaria en temas conocidos.<br>• Estudiante: lee y escribe cartas o notas sobre temas conocidos o predecibles.<br>• Entiende instrucciones y mensajes básicos como catálogos computarizados de biblioteca.<br>• Entiende el significado general de cartas no rutinarias y artículos teóricos dentro de su área de trabajo.<br>• Toma notas en clase, sobre todo si son casi dictadas.<br>• Laboral: entiende el significado general de cartas no rutinarias y artículos teóricos dentro de su área de trabajo.<br>• [Comprende] notas y solicitudes cortas a colegas o conocidos en otra compañía.<br>• Social/turismo: comprende artículos sencillos en periódicos, cartas rutinarias en hoteles, sobre opiniones personales; lee cartas en una variedad de tópicos predecibles relacionados con experiencias personales; entiende opiniones en lenguaje predecible | • General: entiende y expresa opiniones en asuntos abstractos/culturales de manera limitada; [comprende] ofrecimiento de ayuda en temas conocidos.<br>• Estudiante: entiende instrucciones en clases y las tareas que deja el profesor.<br>• Laboral:[comprende] ofrecimiento de ayuda/consejo a clientes en asuntos simples dentro de su área de trabajo.<br>• Social/turismo: expresa opiniones en temas abstractos/culturales de manera limitada y comprende detalles de opinión o inconvenientes. |

Tabla 5.6 Integración de la información de la Tabla 5.1 de acuerdo con las áreas de medición del EXEDII: Comprensión auditiva, Lectura y Gramática, según CILE.

| Organismo consultado | Criterio lectura | Criterio auditiva | Criterio gramática |
|---|---|---|---|
| CILE Etapa intermedia | • Lee con confianza diferentes tipos de textos, distinguiendo temas generales de detalles específicos.<br>• Comprende textos simples narrativos, informativos, descriptivos y argumentativos.<br>• Toma apuntes en clases o seminarios.<br>• [Comprende] ensayos o informes que reflejan un conocimiento básico del tema. | • General: tiene conciencia de registro formal e informal.<br>• Comprende a hablantes nativos de una variedad conocida.<br>• Entiende diferentes puntos de vista en una conversación y puede interactuar expresando hipótesis y eventualidad o imprecisión respecto a hechos del pasado, presente y futuro.<br>• [Comprende] fundamentación de una opinión (argumentos).<br>• Estudiante: puede [comprender] intercambio de opiniones, formulaciones de preguntas y presentaciones breves.<br>• Laboral: puede realizar tareas de oficina, [comprende] intercambio de opiniones en situaciones varias. [Comprende] mensajes escritos y cartas así como contenido general de una conversación cara a cara o en películas, videos, etc. | • Conoce estructuras gramaticales básicas y las aplica con cierta confianza y precisión adecuada a la situación.<br>• Demuestra comprender vocabulario variado en una amplia gama de situaciones. |

Tabla 5.7. Integración de la información de la Tabla 5.1 de acuerdo con las áreas de medición del EXEDII: Comprensión auditiva, Lectura y Gramática según los exámenes ESOL de la Universidad de Cambridge.

| Organismo | Criterio lectura | Criterio auditiva | Criterio gramática |
|---|---|---|---|
| *First Certificate in English (FCE)* | • Lee textos de varios tipos (informativos y general) y entiende la esencia, así como algunos detalles. | • Comprende conversaciones, anuncios, noticias, radio. | • Conocimiento y control del sistema de lenguaje en textos auténticos<br>• Conocimiento léxico y gramatical. |

Para complementar los contenidos de la estructura del lenguaje se seleccionaron como indicadores de los conocimientos mínimos necesarios los indicadores del nivel intermedio de los cuadros de contenidos denominados *SparkCharts™* elaborados por la editorial Barnes & Noble. Esta decisión se deriva de la recomendación de uno de los expertos entrevistados. Los *SparkCharts™* son una especie de atlas o cartografías elaboradas por expertos de la Universidad de Harvard, donde se incluyen los contenidos relevantes y representativos del tema que tratan y están siendo ampliamente utilizados en las instituciones donde se enseña inglés con propósitos específicos. Es de notar que inglés con propósitos específicos se ha convertido en un término técnico, que alude a una metodología de enseñanza de los idiomas, que responde a una necesidad de aprendizaje de grupos de población específicos. Por ejemplo, los estudiantes que tienen necesidad de prepararse para un examen de certificación, o los empresarios que deben aprender a hablar inglés para comunicarse en el ambiente de una maquiladora, o el empresario que requiere aprender, o perfeccionar el manejo de un idioma, para comunicarse en el contexto de las relaciones empresariales internacionales.

Para conocer y evaluar si los *SparkCharts™* tenían la calidad académica necesaria para utilizarse como indicadores de los conocimientos gramaticales mínimos necesarios para el nivel intermedio, se preguntó su opinión sobre el particular a la otra persona entrevistada, quien consideró que eran adecuados.

En el anuncio que aparece en el sitio Web describen a los *SparkCharts™* de la siguiente manera:

> *Imagine if the top student in your course organized the most important points from your textbook or lecture into an easy-to-read, laminated chart that could fit directly into your notebook or binder.*
> *SparkCharts™-created by Harvard students for students everywhere-serve as study companions and reference tools that cover a wide range of subjects, including Business, Math, Science, History, Humanities, Foreign Language, and Writing. Titles like Presentations and Public Speaking, Essays and Term Papers, Resumes and Cover Letters, and Test Prep give you what it takes to find success in college and beyond. Outlines and summaries cover key points, while diagrams and tables make difficult concepts easier to digest.*
>
> Imagine que el estudiante más aventajado de su clase organizara los puntos más importantes del libro de texto en un cuadro que fuera fácil de leer, de aspecto laminado y que cupiera perfectamente en su cuaderno o carpeta. SparkCharts™, creado por estudiantes de Harvard para estudiantes de cualquier escuela sirven como compañeros de estudio y como herramientas de referencia. Cubren una amplia gama de temas como negocios, matemáticas, ciencia, humanidades, lenguas extranjeras y escritura. Los temas como Presentaciones y discursos en público, ensayos y reportes de cursos, cartas de presentación y preparación para exámenes le proporcionan lo que hace falta para lograr éxito en la universidad y en el futuro. Los resúmenes e índices cubren los puntos relevantes, mientras que los diagramas facilitan la comprensión de los conceptos difíciles (Traducción libre de la autora de la tesis).

Por lo anterior, el contenido de Gramática Inglesa *(English Grammar)* de *SparkCharts™* fue utilizado como criterio para contribuir a la conceptualización del DULM, que debe medir el EXEDII. En el Recuadro 5.1 se presentan los contenidos derivados de *SparkCharts™*

Recuadro 5.1 Indicadores de habilidades y conocimientos gramaticales para el nivel intermedio según SparkCharts™

| Indicadores gramaticales de SparkCharts™ |
| --- |
| Sustantivos comunes y propios; singulares y plurales.<br>Forma posesiva (´s).<br>Artículos demostrativos: this, that, these, those y determinativos: a, an, the.<br>Adjetivos: *-less, able/ible, ous, ful, ive.*<br>Comparativos: *-er, est. More than, less than.*<br>Adverbios: *here, there, anywhere, nearbye, indoors.*<br>Action*: Never, once, every, sometimes, next, usually, always.*<br>Intensidad: *very, fairly, rather, quite.*<br>Superlativos: *-st, most/least.*<br>Pronombres personales: *I, you, he, she, it, we, you, they.*<br>Preposiciones: *across, through, before, during, with, without, from, except, instead.*<br>Conjunciones: *and, or, because, although.*<br>Oraciones negativas: *yes/no.*<br>Pedir información*: who, what, where, when, why, how?*<br>Verbos: Presente y pasado simple, continuo, perfecto y futuro simple.<br>Verbos regulares e irregulares; auxiliares: *have, do, be;*<br>Verbos modales: *will, would, can, could, must, may, might, shall, should.* |

### 5.4. Conceptualización del DULM.

La literatura sobre validez de los instrumentos de medición enfatiza la importancia de que el constructo que pretende medir un examen sea claramente definido (por ejemplo Messick, 1993; Popham, 1990). Por ello deben hacerse todos los esfuerzos necesarios para lograr una definición adecuada del constructo a medir. Considerando que la definición del constructo del EXEDII, utilizada para su diseño y construcción es demasiado general para guiar el proceso de validación de contenido, se hizo necesario elaborar una conceptualización de lo que debería medir este examen, a juzgar por los expertos ya que ésta constituye la definición explícita del constructo del EXEDII.

**Cuarta actividad:** construcción de la conceptualización del DULM del EXEDII.

**Propósito**: Redactar la conceptualización del Dominio de Uso de la Lengua Meta, a partir de la integración de los resultados de las tres actividades anteriores.

**Procedimiento**: Se revisaron cuidadosamente las tablas 5.1 a 5.5 con el objeto de tomar sus contenidos como base para describir los criterios de competencia del nivel intermedio en términos de tareas. Estas tareas corresponden a una muestra de las que el sustentante enfrentaría en situaciones de la vida real, por lo que son tareas que deberían estar incluidas en un examen como el EXEDII. No obstante, se descartaron algunas competencias de los estándares internacionales porque abarcan tareas que no están contempladas en el diseño del EXEDII y que, por lo tanto el examen no contiene reactivos para medirlas. Aún así, la conceptualización del dominio de la lengua meta incluye las habilidades y competencias que los jueces manifestaron que deberían formar parte de ese dominio.

En las tablas 5.1 a 5.5 se resaltaron con color amarillo las competencias que a) se refieren a las habilidades receptivas (es decir comprensión auditiva y lectura) puesto que son las que corresponden al diseño del EXEDII, y b) aquellas competencias que a juicio de los expertos panelistas deberían estar incluidas en el DULM, puesto que son las opiniones de ellos las que delimitan el dominio, de acuerdo con el conocimiento que ellos tienen de los sustentantes, sus características y las situaciones en las que los egresados de la UABC típicamente estarían involucrados (Figuras 5.1 y 5.2). El resultado de esta actividad es la conceptualización del DULM, que se presenta en el Recuadro 5.2.

Antes de presentar la conceptualización del dominio que debería medir el EXEDII vale la pena hacer una reflexión acerca de la autenticidad de la evaluación del manejo de un idioma. En todo ejercicio de evaluación se introduce un elemento de artificialidad que permea el proceso desde la intención del evaluador hasta la interpretación de los hallazgos, pasando por los instrumentos de medición, las situaciones en las que se evalúa y la delimitación del evento que se evalúa. Evaluar las habilidades para el manejo del lenguaje implica una separación conceptual de los componentes de la competencia lingüística a medir, separación que necesariamente es arbitraria e incompleta. Pero aún en las ciencias naturales, en las que el objeto de estudio se presta a delimitaciones más claras, en ocasiones se requiere recurrir a ejercicios de disección, en los que se eliminan los componentes que no se consideran esenciales para poder estudiar el objeto con mayor objetividad. Por ello, en estudios como el que se reporta en este capítulo se hacen esfuerzos por delimitar el componente receptivo de la situación de uso de la lengua para poder evaluar las cualidades del instrumento de medición. Los hallazgos deben entenderse como tales, aún a sabiendas de que se está violando la autenticidad de la situación evaluada.

Recuadro 5.2. Conceptualización del Dominio de Uso de la Lengua Meta (DULM).

El dominio de uso de la lengua de los sustentantes del EXEDII se caracteriza por:

- La demostración de habilidades comunicativas en situaciones generales que pueden incluir tareas en ambientes laborales, escolares, turísticos y sociales.
- En esas situaciones deben ser capaces de mantener una conversación sencilla con interlocutores angloparlantes, ya sean nativos o no nativos en situaciones cara a cara o en pequeños grupos, cuando los interlocutores hablan relativamente despacio y muestran paciencia para escucharlos.
- En este tipo de situaciones los sustentantes pueden comprender el tema general del que se está hablando, así como algunos detalles cuyo significado e intencionalidad derivan del contexto, apoyándose en sus habilidades en la lengua materna.
- También pueden comprender asuntos abstractos o culturales que les son familiares, particularmente en situaciones de intercambio comunicativo con sus pares, o con personas conocidas donde se expresan diferentes puntos de vista, opiniones y argumentos con respecto al pasado, presente y futuro siempre que se trate de temas conocidos, pudiendo formular preguntas o realizar intervenciones breves.
- Comprender preguntas de un interlocutor como la solicitud de información sobre su persona, familia y actividades diarias, así como de otros temas que les son conocidos.
- En situaciones como la vida escolar son capaces de entender las instrucciones generales que se dan en la clase y las tareas que deja el profesor.
- Tratándose de situaciones laborales son capaces de captar la solicitud de un cliente y ofrecer ayuda o consejo en asuntos simples y dentro de su área de trabajo.
- En situaciones de comunicación no interactiva son capaces de entender la idea general de estímulos tales como anuncios, noticias, radio, televisión y cine o letras de canciones.
- En las diferentes situaciones de comunicación son capaces de discriminar el registro formal o informal (P.ej.el lenguaje formal de una conferencia o el coloquial entre amigos).
- Las habilidades de comprensión lectora de los sustentantes del EXEDII incluyen la comprensión de la idea general y detalles derivados del contexto en una variedad de textos como por ejemplo informativos, de divulgación, textos de naturaleza descriptiva, narrativa y discursiva con información de rutina y/o de divulgación.
- Si se trata de textos con información no rutinaria, comprenden los relativos a temas que les resultan conocidos como académicos de su área, pasatiempos, deportes.
- Pueden comprender cartas o notas sobre temas conocidos o predecibles, así como instrucciones y mensajes básicos (P. ej. catálogos computarizados de bibliotecas).
- Comprenden también textos sencillos en periódicos, cartas rutinarias en hoteles y servicios en general, textos sobre opiniones personales y pueden comprender notas y solicitudes cortas con colegas o conocidos sobre temas de su área de trabajo, personales o escolares.
- En el ambiente escolar entienden la idea general de ensayos, artículos o reportes sobre temas técnicos o académicos siempre que se refieran a su área de estudio o trabajo.
- Pueden entender detalles específicos de temáticas que les son conocidas.
- En cuanto a la estructura del idioma, conocen reglas gramaticales que le permiten comunicarse si se trata de eventos del pasado, presente y futuro, aunque pueden cometer errores de sintaxis, precisión, alcance y pronunciación. Su vocabulario es limitado, pero suficiente para la comunicación de temas conocidos.

## 5.5. Comparación del contenido del EXEDII contra el criterio.

De acuerdo con la definición de validez de contenido de los estándares internacionales (*Standards for Educational and Psychological Testing,* 1985 página 10) "la evidencia relativa al contenido demuestra el grado en el que la muestra de reactivos, tareas o preguntas de una prueba es representativa de un universo o dominio de contenido definido." Consecuentemente, la tercera pregunta de investigación que pretende ser respondida en este capítulo dice:

> ¿Los reactivos del EXEDII constituyen una instancia del uso del inglés de los estudiantes que egresan de la UABC? y si ese es el caso ¿de qué manera estos datos aportan evidencias de validez de contenido y de constructo de la prueba?

La postura que se ha tomado en la presente investigación plantea que la evaluación del manejo que los individuos tienen de un idioma debe implicar una demostración de la correspondencia entre el uso de la lengua en el mundo real y la situación de prueba. Para estar en posibilidad de realizar esa demostración es necesario asegurar que el instrumento con el que se evalúa representa una instancia del uso de la lengua meta, en la que se toman en cuenta las características esenciales de los sustentantes, de las tareas y de las situaciones en las que se daría un uso determinado de la lengua meta (Bachman y Palmer, 1996).

En este apartado se describen las actividades realizadas para recabar evidencias acerca de la correspondencia entre la prueba y lo que pretende medir, a través de juzgar la relevancia, representatividad y pertinencia de los reactivos, como instancias del DULM. Así mismo, se valoran las cualidades de autenticidad e interactividad de las tareas descritas en los reactivos del EXEDII, por considerar que estas dos cualidades son indispensables para interpretar el concepto de "el mundo real" aludido en la definición de Bachman y Palmer. Finalmente se investiga el posible sesgo de los reactivos.

La metodología incluye la participación de un grupo de expertos, quienes revisaron y evaluaron los reactivos del EXEDII comparándolos contra la conceptualización del DULM, apegándose a los criterios evaluativos tal y como se definen en las Escalas de Validación, de las cuales se presentan enseguida los fragmentos correspondientes, pero pueden consultarse en extenso en el Anexo 5.

Recuadro 5.3. Definición de los criterios evaluativos para los Nodos tal como aparecen en las Escalas de Validación.

| Criterios evaluativos para los Nodos de las subescalas de Comprensión auditiva, Gramática y Lectura. |
|---|
| De acuerdo con la conceptualización del DULM, EL NODO DEL REACTIVO ES... |
| 1. Relevante para el uso del lenguaje que pretende medir EXEDII.<br>Es decir, el nodo del reactivo hace referencia a una tarea que muy probablemente el sustentante del EXEDII enfrentaría en el mundo real. |
| 2. Representativo del uso del lenguaje que pretende medir EXEDII.<br>Es decir, el nodo del reactivo hace referencia a una tarea que típicamente enfrentaría el sustentante del EXEDII en el mundo real. |
| 3. Pertinente para el constructo a medir.<br>Es decir, el nodo del reactivo hace referencia a una tarea que está incluida entre las tareas que el sustentante del EXEDII enfrentaría en el mundo real. |
| 4. Definido de manera clara y sin ambigüedad.<br>Es decir, el nodo del reactivo expresa cabalmente el tipo de tarea que el sustentante del EXEDII tendría que enfrentar en el mundo real. |

Para la valoración de las características de los reactivos de las tres subescalas los criterios se ilustran en el Recuadro 5.4:

Recuadro 5.4. Definición de los criterios evaluativos para las Características de los reactivos, tal como aparecen en las Escalas de Validación.

| Criterios evaluativos para las Características de los reactivos<br>de las subescalas de Comprensión auditiva, Gramática y Lectura. |
|---|
| De acuerdo con la conceptualización del DULM, las características del reactivo... |
| 5. Representan una situación de ese uso del lenguaje.<br>Es decir, la tarea de este reactivo se parece en sus características esenciales, a las tareas que el sustentante enfrentaría en el mundo real. |
| 6. Suponen el nivel apropiado de conocimiento.<br>Es decir, la tarea de este reactivo implica una tarea que el sustentante puede resolver con su conocimiento y habilidades en el uso del lenguaje. |
| 7. Reflejan las características de los sustentantes.<br>Es decir, la tarea de este reactivo es adecuada para el sustentante del EXEDII, que es típicamente un estudiante que egresa de alguna de las carreras de la UABC . |
| 8. Reflejan el contexto cultural de los sustentantes.<br>Es decir, la tarea de este reactivo toma en cuenta el ambiente, las necesidades y características en las que típicamente se desenvuelve el sustentante. |
| 9. Provocan sesgo.<br>Es decir, la redacción de este reactivo favorece la respuesta diferencial de los sustentantes por motivos diferentes a su habilidad, como por ejemplo su clase social, su género, su edad, etc. |

Los criterios evaluativos son específicos para cada subescala por lo que los siguientes tres recuadros ilustran las respectivas definiciones:

Recuadro 5.5. Definición de los criterios evaluativos para las habilidades/competencias de los reactivos, tal como aparecen en la Escala de Validación de Comprensión auditiva.

| Definición de los criterios evaluativos para las habilidades/competencias de la subescala de Comprensión auditiva. |
|---|
| El REACTIVO MIDE LAS SIGUIENTES COMPETENCIAS INCLUIDAS EN EL DULM |
| 10. Discriminar registro formal/informal. |
| Es decir, el sustentante puede diferenciar cuándo debe usar lenguaje formal o informal (por ejemplo, con profesores, compañeros, jefes, etc.). |
| 11. Discriminar opiniones de interlocutor(es). |
| Es decir, el sustentante puede diferenciar cuando alguien emite su opinión en una situación de uso del lenguaje. |
| 12. Comprender preguntas de un interlocutor. |
| Es decir, el sustentante puede entender el tema sobre el que su interlocutor lo cuestiona, o bien cuando alguien es cuestionado en una situación de uso del lenguaje real, o en un texto escrito. |
| 13. Comprender a hablantes nativos de variedad conocida, si le hablan despacio y claramente. |
| Es decir, el sustentante puede entender el tema al que se refiere su interlocutor en una situación de uso del lenguaje real, o en un texto escrito, siempre que se trate del tipo de inglés con el que está familiarizado el sustentante típico del EXEDII, que corresponde al del sur de California, aunque pudiera ser otro. |
| 14. Comprender inconvenientes (I don't like...). |
| Es decir, el sustentante puede entender en una situación de uso del lenguaje real, o en un texto escrito, cuando algo no es deseable, o le causa algún tipo de conflicto. |
| 15. Comprender instrucciones en clase y tareas que deja el profesor. |
| Es decir, el sustentante puede entender en una situación de uso del lenguaje real, o en un texto escrito, cuando un profesor solicita a los estudiantes que realicen alguna actividad relacionada con la vida académica. |
| 16. Comprender diferentes puntos de vista en tiempo pasado, presente y/o futuro sobre temas conocidos. |
| Es decir, el sustentante puede entender, en una situación de uso del lenguaje real, o en un texto escrito, cuando alguna acción sucede en cualquiera de estos tres tiempos gramaticales. |
| 17. Comprender la solicitud de información sobre sí mismo. |
| Es decir, el sustentante puede entender en una situación de uso del lenguaje real, o en un texto escrito, cuando alguien pide o da información de tipo personal (nombre, edad, gustos y preferencias, opiniones, problemas, etc.). |
| 18. Comprender la solicitud de ayuda o consejo en asuntos simples y dentro de su área de conocimiento. |
| Es decir, el sustentante puede comprender la situación en la que alguien colabora con, o auxilia a alguien, en algún asunto que le es familiar. |
| 19. Comprender el intercambio de opiniones, formulación de preguntas o conversaciones de sus pares. |
| Es decir, el sustentante puede entender las conversaciones entre personas, en una situación de uso del lenguaje real, o en un texto escrito, cuando se realizan preguntas, se intercambian opiniones o puntos de vista. |
| 20. Comprender la interacción en temas abstractos o culturales sobre temáticas conocidas. |
| Es decir, el sustentante puede entender en una situación de uso del lenguaje real, o en un texto escrito, interacciones de comunicación en las que se tocan temas abstractos o de índole cultural, siempre que se refieran a situaciones conocidas para el sustentante. |
| 21. Comprender la argumentación o planteamiento de opiniones personales. |
| Es decir, el sustentante puede entender en una situación de uso del lenguaje real, o en un texto escrito, cuando alguien ofrece argumentos que sustenten una opinión personal, o punto de vista. |

En el siguiente recuadro se presentan los criterios evaluativos de las habilidades/competencias para la subescala de Gramática:

Recuadro 5.6. Definición de los criterios evaluativos para las habilidades/competencias de los reactivos, tal como aparecen en la Escala de Validación de Gramática.

| Definición de los criterios evaluativos para las habilidades/competencias de la subescala de Gramática. |
|---|
| EL REACTIVO MIDE LAS SIGUIENTES COMPETENCIAS INCLUIDAS EN EL DULM<br>10. Uso de sustantivos comunes y propios: singular y plural.<br>11. Forma posesiva: ´s<br>12. Artículos demostrativos: *this, that, these, those* y determinativos: *a, an, the.*<br>13. Adjetivos: less, able/ible, ous, ful, ive.<br>14. Comparativos: *er, est; more than, less than*<br>15. Adverbios: *here, there, anywhere, nearbye, indoors; Now, then, later, early, tomorrow, next year, already, not yet, still; Never, once, every week, sometimes, next, usually, always. Carefully, slowly; Very, fairly, rather, quite.*<br>16. Superlativos: *-st, most/least.*<br>17. Pronombres personales: *I, you, he, she, it, we, you, they.*<br>18. Preposiciones: *across, through, before, during, with, without, from, except, instead.*<br>19. Conjunciones: and, or, because, although.<br>20. Oraciones negativas: *I am not, I do not, I have not.*<br>21. Pedir información: *who, what, where, when, why, how?*<br>22. Verbos: regulares e irregulares.<br>23. Tiempos gramaticales: presente y pasado simple, continuo, perfecto y futuro simple.<br>24. Verbos auxiliares: *have, do, be.*<br>25. Verbos modales: *will, would, can, could, must, may, might, shall, should.* |

Para complementar la información sobre los criterios a los que se apegaron los panelistas que evaluaron los reactivos del EXEDII en el siguiente recuadro se presentan los criterios evaluativos de las habilidades/competencias para la subescala de Lectura:

Recuadro 5.7. Definición de los criterios evaluativos para las habilidades/competencias de los reactivos, tal como aparecen en la Escala de Validación de Lectura.

| Criterios evaluativos para las habilidades/competencias de la subescala de Lectura. |
|---|
| EI REACTIVO MIDE LAS SIGUIENTES COMPETENCIAS INCLUIDAS EN EL DULM |
| 10. Comprende textos con información de rutina y artículos de divulgación. |
| 11. Comprende textos con información no rutinaria en temas conocidos. |
| 12. Comprende cartas o notas sobre temas conocidos o predecibles. |
| 13. Comprende instrucciones y mensajes básicos como catálogos computarizados de biblioteca. |
| 14. Comprende notas y solicitudes cortas a colegas o conocidos en temas de su área de trabajo. |
| 15. Comprende textos sencillos en periódicos, cartas rutinarias en hoteles y servicios en general. |
| 16. Comprende textos sobre opiniones personales. |
| 17. Comprende ensayos o artículos/reportes sobre temas técnicos/académicos de su área de estudio/trabajo. |
| 18. Comprende notas en clase, apoyándose en habilidades en lengua materna. |

**Conformación del Panel de expertos.**

Se conformaron tres paneles de expertos, de acuerdo con las tres subescalas que constituyen el EXEDII, es decir hubo un panel de Comprensión auditiva, otro de Lectura y un tercero de Gramática.

**Propósitos**: Establecer las condiciones para que los panelistas evaluaran: a) la representatividad, relevancia y pertinencia de las tareas; b) la autenticidad e interactividad de cada reactivo; c) la pertinencia de las competencias o habilidades medidas, así como el posible sesgo derivado de las características de los reactivos.

**Procedimiento**: se utilizó la metodología denominada juicio post facto (Popham, 1990:98) o "jueceo", la cual consiste en reunir a un grupo de expertos y solicitarles que analicen las características de un instrumento de evaluación educativa determinado y emitan su juicio, de acuerdo con algún criterio

recomendado. El jueceo también contempla la discusión grupal para permitir que se genere información que no sería evidente con la calificación individual de los aspectos evaluados.

**Participantes:** un total de 12 expertos en docencia y/o en la enseñanza del inglés. En la Tabla 5.8 se observa que seis expertos conformaron el Panel de Comprensión Auditiva:

Tabla 5.8. Conformación del panel de expertos para la evaluación de los reactivos del Comprensión auditiva.

| PANEL DE COMPRENSION AUDITIVA | | |
|---|---|---|
| Perfil Deseado | Características relevantes | Área académica |
| LAE | Docente | Económico-Administrativa |
| Médico | Docente/Coordinador Medicina | Químico-Biológicas/ Salud |
| Ingeniero | Docente/Coordinadora | Ingeniería |
| Lic. Ciencias Comunicación | Docente Facultad Ciencias Administrativas y Sociales | Ciencias Sociales |
| Docente de inglés | Docente/Funcionario público experto en TOEFL | Inglés |
| Docente de ingles | Docente Facultad Idiomas | Inglés |

De manera similar el Panel de Lectura estuvo conformado por seis docentes de diferentes carreras de la UABC. La Tabla 5.9 muestra la conformación de este panel:

Tabla 5.9. Conformación del panel de expertos para la evaluación de los reactivos de Lectura del EXEDII.

| PANEL DE LECTURA | | |
|---|---|---|
| Perfil Deseado | Características relevantes | Área académica |
| Contador Público | Docente Facultad Ciencias Administrativas y Sociales, UABC. | Económico-Administrativa |
| Biólogo | Docente en la Facultad Ciencias | Químico-Biológicas/ Salud |
| Lic. Ciencias computacionales | Docente Facultad Ingeniería | Ingeniería |
| Socióloga | Docente/ Coordinadora Sociología Facultad Ciencias Administrativas y Sociales. | Ciencias Sociales |
| Docente de inglés | Docente Facultad Idiomas | Inglés |
| Docente de inglés | Docente en la Escuela Normal estatal | Inglés |

El Panel de Gramática estuvo conformado por los cuatro docentes de inglés que en una actividad previa participaron en los paneles de Comprensión auditiva y de Lectura. La Tabla 5.10 muestra la conformación de este panel:

Tabla 5.10. Conformación del panel de expertos para la evaluación de los reactivos de Gramática del EXEDII.

| PANEL DE GRAMATICA | | |
|---|---|---|
| Perfil Deseado | Características relevantes | Área académica |
| Docente de inglés | Docente Facultad Idiomas | Inglés |
| Docente de inglés | Docente Facultad Idiomas | Inglés |
| Docente de inglés | Docente Facultad Idiomas/Funcionario público experto en TOEFL. | Inglés |
| Docente de inglés | Docente en la Escuela Normal estatal. | Inglés |

Características de los expertos. Se buscaron dos tipos de expertos: a) panelistas que tuvieran experiencia en docencia en la UABC, independientemente de su manejo del inglés ya que se buscaba que ellos valoraran los aspectos de autenticidad e interactividad de los reactivos es decir, si éstos representan situaciones y tareas en las que los sustentantes podrían verse involucrados como estudiantes, egresados y profesionistas en la etapa inicial de su vida profesional. Los profesores de las carreras de la UABC tienen conocimiento de las características de los estudiantes de esa universidad, de su contexto sociocultural y de las demandas que tienen como profesionales en el campo laboral, conocimiento que es indispensable para juzgar las características mencionadas de los reactivos de comprensión auditiva y lectura; b) panelistas que tuvieran experiencia en la enseñanza y evaluación del inglés, que estuvieran en condiciones de juzgar los aspectos evaluados en los reactivos de auditiva y lectura y aportar su opinión como expertos en inglés, en los casos en los que no hubiese acuerdos en la calificación. Por las mismas características se reunió a los mismos cuatro expertos en inglés para conformar el panel que requería que los jueces fueran expertos en esa materia, es decir el panel de Gramática. La organización y conducción de la sesión estuvo a cargo de la autora de este trabajo de tesis.

**Selección de la muestra.**

La selección de la muestra de participantes fue intenciona, por lo que se invitó a un número de docentes que tuvieran el perfil requerido para participar en este panel, de acuerdo con las características detalladas en el apartado anterior, las cuales cubren: a) familiaridad con el contexto cultural de los sustentantes y/o b) conocimiento de las características de las tareas del uso de la lengua

meta y/o c) familiaridad con las características de los sustentantes, que son factores que influyen de manera determinante en la utilidad del instrumento (Bachman y Palmer,1966).

**Invitación a participar**

Una vez diseñada la estructura de los paneles, se invitó personalmente, por teléfono y/o por correo electrónico a los posibles panelistas y se les explicó el propósito de la actividad. Se ofreció una remuneración por su colaboración, así como un reconocimiento por escrito. A los profesores que aceptaron la invitación se les citó en el salón de evaluación del Sistema de Información Académica de la UABC, campus Ensenada, para darles oportunidad de que resolvieran el EXEDII en las mismas condiciones en que lo presentan los sustentantes típicos y pudieran así percatarse de las competencias lingüísticas que se ponen en juego al responder el EXEDII. Cuando todos los panelistas terminaron, se trasladaron al Salón Virtual del Instituto de investigación y Desarrollo Educativo (IIDE) de la UABC. Una vez que estuvieron presentes todos los panelistas se les dio una breve introducción al tema de la validación de instrumentos, al concepto de Uso de la Lengua Meta y se explicó la mecánica de la sesión, así como todas las instrucciones necesarias para evaluar los reactivos de la sección del EXEDII que les correspondía; se entregaron los instrumentos y se inició la actividad.

**Procedimiento**

La evaluación o "jueceo" se dio en dos tiempos: los paneles de Comprensión auditiva y de Lectura trabajaron primero; cuando terminaron de trabajar se retiraron los profesores de las materias curriculares y se quedaron los expertos en inglés quienes conformaron el panel de Gramática. Se tomó la decisión de "reciclar" los paneles por las siguientes razones: 1) los expertos en inglés tuvieron la oportunidad de conocer la percepción de los docentes acerca del contexto cultural de los estudiantes porque participaron en uno de los dos paneles anteriores 2) la discusión grupal y el logro de consensos se facilita con paneles de tamaño mediano (Álvarez Gayou, 2005), pero de extensión suficiente para el "jueceo" de contenidos (Popham, 1990).

**Instrumentos.** Se entregó a cada panelista una carpeta de trabajo que contenía los siguientes instrumentos:

- En un fólder de pasta dura, con la carátula transparente se introdujo una tarjeta tamaño carta, de color amarillo, la cual tenía impresa la Conceptualización del Dominio de Uso de la Lengua Meta (ver Conceptualización del DULM en el apartado 6.2.2) misma que

se les pidió que leyeran cuidadosamente. Cuando la terminaron de leer se les pidió que la mantuvieran disponible en todo momento. Así mismo se les pidió que la consultaran en caso de necesitar recordar detalles de las competencias, o de las tareas de ese dominio.

- Un Cuaderno de reactivos, tamaño media carta, donde se incluyeron todos los reactivos (por área de medición), uno por página.
- Escala de Validación, tamaño oficio, que contenía una tabla de doble entrada conteniendo en el plano vertical, los aspectos a evaluar y en el plano horizontal, celdillas que corresponden a los números de los reactivos de cada sección evaluada. Los aspectos a evaluar están redactados en una serie de aseveraciones que se incluyen en los recuadros 5.1 a 5.6, arriba.

Para evaluar cada reactivo, los panelistas leyeron cada nodo en el Cuaderno de reactivos, después leyeron cada aseveración y su respectiva explicación y después de reflexionar su respuesta anotaron un 1 ó un 0 en la celdilla correspondiente al reactivo, en la Escala de Validación. El número 1 significaba "totalmente de acuerdo con la aseveración" y el 0 significaba "totalmente en desacuerdo con la aseveración".

Al final de la Escala de Validación se incluyó espacio para que los jueces justificaran las respuestas marcadas con 0 (en desacuerdo).

Al terminar de calificar todos los reactivos, los panelistas compararon sus respuestas con los demás compañeros de su Panel y cuando hubo discrepancias, discutieron y ofrecieron sus argumentos para tales respuestas. En algunas ocasiones, la discusión llevó a algunos jueces a cambiar su respuesta, por considerar válidos los argumentos de sus colegas, pero en algunas otras ocasiones, la discrepancia prevaleció y quedó anotada en la Escala de Validación, junto con su respectiva justificación.

### 5.5. Resultados

Se construyó una base de datos con las respuestas de los panelistas que compararon los reactivos del EXEDII con el DULM para las tres escalas del examen. En las columnas se incluyeron los aspectos evaluados y en los renglones los reactivos, ordenados por jueces. Con ayuda del paquete

estadístico *Statistics Program for the Social Sciences (SPSS)* se obtuvieron las frecuencias relativas para las respuestas de cada panelista, así como los porcentajes derivados de ellas.

**Análisis de la apreciación de los Nodos y características de los reactivos.**

Apelando a las recomendaciones de la literatura (Popham, 1990; Heaton, 1988) se tomó el criterio que establece que los acuerdos que no alcancen el 75% se consideran indicadores de que el reactivo merece una revisión más cuidadosa en el aspecto evaluado que obtenga ese resultado. A continuación se ofrece el análisis de la subescala de Comprensión auditiva en la Tabla 5.11 (en el Anexo 5.1 se pueden consultar las tablas con los datos completos para las tres subescalas).

Tabla 5.11. Acuerdos entre los panelistas respecto de los aspectos evaluados de los Nodos y Características de los reactivos, expresados en frecuencias relativas y porcentajes. Subescala: Comprensión auditiva.

| | NODOS | | | | | | | | CARACTERISTICAS DE LOS REACTIVOS | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Relevante | | Representativo | | Pertinente | | Claramente descrito | | Similar a la realidad | | Nivel apropiado | | Refleja características de sustentantes | | Refleja contexto cultural de sustentantes | | Provoca sesgo | |
| | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % |
| 14 | | | | | | | | | | | 4/6 | 66.7 | | | | | | |
| 17 | | | | | | | 4/6 | 66.7 | 4/6 | 66.7 | 5/6 | 83.3 | 4/6 | 66.7 | 4/6 | 66.7 | | |
| 18 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | | | | | | | 4/6 | 66.7 | | |
| 24 | | | | | | | | | | | | | | | 4/6 | 66.7 | | |
| 25 | | | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | | | | | 4/6 | 66.7 | 4/6 | 66.7 | | |
| 26 | 3/6 | 50 | | | 3/6 | 50 | 2/6 | 33.3 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | | |
| 27 | | | | | | | | | | | | | | | 4/6 | 66.7 | | |
| 28 | | | | | | | | | | | | | | | 4/6 | 66.7 | | |
| To-tal | 3/32 | 9.4 | 3/32 | 9.4 | 3/32 | 9.4 | 4/32 | 12.5 | 2/32 | 6.2 | 3/32 | 9.4 | 3/32 | 9.4 | 7/32 | 21.8 | - | - |

FR: frecuencia relativa con N=6 jueces. Número de reactivos de la subescala: 32
Las celdas en blanco representan consenso de los jueces respecto a la respuesta "De acuerdo" con la aseveración que define cada aspecto evaluado.
Cuando los jueces no llegaron a consensos y el porcentaje de acuerdo es menor al 75%, se expresan las frecuencias relativas correspondientes (4/6, 3/6, 2/6 ó 1/6) y su valor en porcentajes.

El análisis cuantitativo de la apreciación que los jueces hicieron de los Nodos y de las Características de los reactivos muestra que el 9.4% de ellos no se consideró Relevante, Representativo y Pertinente y el 12.5% no está definido con claridad. A propósito de las características de los reactivos el 6.2% se perciben como alejados de lo que sería una situación Similar a la realidad; el 9.4% no corresponde al Nivel apropiado y no reflejan las características de los sustentantes; el 21.8% no Refleja el contexto cultural de los sustentantes y no existe indicación significativa de problemas de sesgo.

La Tabla 5.12. expone los resultados para la subescala de Gramática donde se puede observar que todos los aspectos evaluados para los Nodos y para las Características de los reactivos fueron calificadas como adecuadas. No obstante, de los 34 reactivos de esta subescala el 29.4% no reflejan el contexto cultural de los sustentantes, a decir de los jueces.

Tabla 5.12. Acuerdos entre los panelistas respecto de los aspectos evaluados de los Nodos y Características de los reactivos, expresados en frecuencias relativas y porcentajes. Subescala: Gramática

| | NODOS | | | | | | | | | | CARACTERISTICAS | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Relevante | | Representativo | | Pertinente | | Claramente descrito | | Similar a la realidad | | Nivel apropiado | | Refleja características de los sustentantes | | Refleja contexto cultural de sustentante | | Provoca sesgo | |
| | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % |
| R36 | | | | | | | | | | | | | | | 2/4 | 50 | | |
| R41 | | | | | | | | | | | | | | | 3/4 | 75 | | |
| R47 | | | | | | | | | | | | | | | 3/4 | 75 | | |
| R48 | | | | | | | | | | | | | | | 1/4 | 25 | | |
| R51 | | | | | | | | | | | | | | | 3/4 | 75 | | |
| R52 | | | | | | | | | | | | | | | 3/4 | 75 | | |
| R53 | | | | | | | | | | | | | | | 3/4 | 75 | | |
| R54 | | | | | | | | | | | | | | | 3/4 | 75 | | |
| R55 | | | | | | | | | | | | | | | 3/4 | 75 | | |
| R58 | | | | | | | | | | | | | | | 3/4 | 75 | | |
| Total | - | - | - | - | - | - | - | - | - | - | - | - | - | | 10/34 | 29.4 | - | - |

FR: frecuencia relativa con N=4jueces. Número de reactivos de la subescala: 34
Las celdas en blanco representan consenso de los jueces respecto a la respuesta "De acuerdo" con la aseveración que define cada aspecto evaluado.
Cuando los jueces no llegaron a consensos y el porcentaje de acuerdo es menor al 75%, se expresan las frecuencias relativas correspondientes (2/4 ó 1/4) y su valor en porcentajes.

Tabla 5.13. Acuerdos entre los panelistas respecto de los aspectos evaluados de los Nodos y Características de los reactivos, expresados en frecuencias relativas y porcentajes. Subescala: Lectura

| | NODOS | | | | | | | | CARACTERISTICAS DE LOS REACTIVOS | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Relevante | | Representativo | | Pertinente | | Claramente descrito | | Similar a la realidad | | Nivel apropiado | | Refleja características de sustentantes | | Refleja contexto cultural de sustentantes | | Provoca sesgo | |
| | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % |
| R75 | | | | | | | | | | | | | | | | | | |
| R77 | | | | | | | | | | | | | | | | | 3/6 | 50 |
| R78 | | | 4/6 | 66.7 | | | | | | | 4/6 | 66.7 | | | 4/6 | 66.7 | | |
| R82 | | | | | | | | | | | 4/6 | 66.7 | | | 4/6 | 66.7 | 3/6 | 50 |
| R83 | | | | | | | | | | | | | | | | | 3/6 | 50 |
| R85 | | | | | | | 4/6 | 66.7 | | | | | | | 3/6 | 50 | 3/6 | 50 |
| R86 | | | | | | | | | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | 3/6 | 50 | 3/6 | 50 |
| R89 | | | 4/6 | 66.7 | | | | | | | | | | | 4/6 | 66.7 | 3/6 | 50 |
| R90 | | | | | | | | | | | 4/6 | 66.7 | | | 4/6 | 66.7 | 4/6 | 66.7 |
| R91 | | | | | | | | | | | | | | | | | 3/6 | 50 |
| R94 | | | | | | | | | | | | | | | 4/6 | 66.7 | 3/6 | 50 |
| R95 | | | | | | | | | | | | | | | 4/6 | 66.7 | | |
| R96 | | | | | | | | | | | | | | | 4/6 | 66.7 | | |
| Total | - | - | 2/34 | 5.9 | - | - | 1/34 | 2.9 | 1/34 | 2.9 | 4/34 | 11.8 | 1/34 | 2.9 | 9/34 | 26.5 | 9/34 | 26.5 |

FR: frecuencia relativa con N=6 jueces. Número de reactivos de la subescala: 34

Las celdas en blanco representan consenso de los jueces respecto a la respuesta "De acuerdo" con la aseveración que define cada aspecto evaluado.

Cuando los jueces no llegaron a consensos y el porcentaje de acuerdo es menor al 75%, se expresan las frecuencias relativas correspondientes (4/6, 3/6, 2/6 ó 1/6) y su valor en porcentajes.

Finalmente, el juicio de los expertos con respecto a los Nodos y las Características de los reactivos de la subescala de Lectura se presentan en la Tabla 5.13. Se observa que el 5.9% de los Nodos no se consideran representativos del dominio evaluado; el 2.9% no se perciben como Claramente descritos, Similares a la realidad, ni reflejan las Características de los sustentantes. Por otra parte el 11.8% no se consideraron del Nivel apropiado y el 26.5 no refleja el contexto cultural de los sustentantes.

**Análisis de la apreciación de las habilidades/competencias.**

El análisis de la evaluación de las habilidades/competencias que supone cada reactivo del EXEDII se realizó de la misma manera que se procedió con los Nodos y las Características de los reactivos. Como se ha venido exponiendo a lo largo de este capítulo, los nodos de los reactivos del EXEDII fueron diseñados para medir habilidades, pero los estándares internacionales consultados definen en términos de competencias la combinación de conocimientos, habilidades y actitudes que estaría demostrando una persona que responde a una situación particular de la vida real, en la que su manejo del lenguaje queda incluido en un dominio, denominado Dominio de Uso de la Lengua. También se explicó que ese dominio se acota aún más para adecuarlo al de una población determinada que es la que contesta el examen que está siendo evaluado y se denomina Dominio de Uso de la Lengua Meta (DULM).

Lo anterior se planteó nuevamente para enlazar la lógica que está detrás de las tablas que siguen:

Tabla 5.14. Resultados de las valoraciones de los jueces con respecto a la alineación de los Nodos de la subescala de Comprensión auditiva con los contenidos del DULM.

| % | Registro | Opinión/ Hecho | Pregón-ta | Nativos | Inconve -nientes | Instruc/ tareas | Puntos de vista | Info perso-nal | Solicita ayuda | Inter/ Opinión | Abstrac/ cultural | Argu-mento |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 0 | 2 | 1 | 0 | 4 | 3 | 5 | 1 | 1 | 2 | 3 | 8 |
| 75 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -75% | 1 | 2 | 0 | 0 | 4 | 3 | 1 | 0 | 0 | 2 | 2 | 6 |
| 0 | 31 | 28 | 31 | 32 | 24 | 26 | 26 | 31 | 31 | 28 | 27 | 18 |
| Total | 1 | 4 | 1 | 0 | 8 | 6 | 6 | 1 | 1 | 4 | 5 | 14 |

La Tabla 5.14 muestra las competencias que se incluyen en el DULM y que, de acuerdo con las respuestas de los expertos se relacionan con los Nodos de la subescala de Comprensión auditiva del EXEDII. Los datos que aparecen en los renglones de la tabla informan sobre el número de reactivos que se alinean con las competencias del DULM. Así puede observarse en el último renglón que un reactivo apunta a la competencia de discriminar el Registro del Uso de la lengua. Lo mismo puede decirse de las competencias de Formular preguntas, ofrecer información personal y solicitar ayuda. Cuatro reactivos inciden en la competencia relativa a la comprensión de estímulos aurales en los que alguien emite opiniones. Cinco reactivos pueden detectar la comprensión del discurso hablado que tiene que ver con aspectos culturales, relativamente abstractos. Dos reactivos están orientados a la comprensión de instrucciones y asignaciones de tareas y al planteamiento de puntos de vista. Ocho reactivos se relacionan con la comprensión de expresiones de algún tipo de inconveniente y finalmente, catorce reactivos son sensibles a la presentación de argumentos o razones para afirmar algo. Solamente la competencia que establece la capacidad, en sentido genérico de interactuar y comprender el discurso de hablantes nativos de la lengua inglesa no contempla a ningún reactivo del EXEDII.

En la misma tabla, pero en el segundo renglón se observa que un reactivo de las competencias Comprensión de preguntas, Información personal y Solicitud de ayuda lograron consenso de los jueces acerca de que el Nodo mide la habilidad para la que fue diseñado. Las competencias que se relacionan con la discriminación de Opinión logran el consenso en cuatro

reactivos. También lograron consenso tres reactivos de las competencias de Comprensión de instrucciones o tareas escolares y el discurso de temas que, aunque traten de asuntos conocidos para el oyente, incluyen información relativamente abstracta o de índole cultural. La expresión de Inconvenientes incluye el consenso de los jueces en cuatro reactivos y la de Puntos de vista en cinco. Ningún reactivo logra consenso al respecto de la competencia/habilidad de discriminar el registro, aunque un reactivo obtuvo menos del 75% de acuerdo y por esa razón se incluye en el resultado final que aparece en el último renglón de la tabla.

Enseguida se muestran en la Tabla 5.15 los resultados correspondientes a la apreciación de las competencias/habilidades de la subescala de Gramática. Esta tabla se debe leer de la misma manera que la anterior, por lo que observando el último renglón se puede verificar el total de reactivos que se alinean a cada uno de los temas que deberían conocer los sustentantes del EXEDII, desde el punto de vista de la estructura de la lengua inglesa. Así, puede notarse que los jueces opinan que un reactivo detecta el uso de conjunciones, así como el uso de negativos. Dos reactivos se relacionan con el uso de palabras interrogativas como *who* y *what* en frases interrogativas y otros dos con preposiciones. Tres reactivos valoran el uso de adjetivos terminados en "*ble*", "*ous*" y otras similares. Cuatro reactivos miden el uso de superlativos, cinco el uso de adjetivos comparativos, seis el de adverbios y nueve el de pronombres personales. Un número mayor de reactivos están relacionados con el conocimiento de verbos auxiliares, que son 14; 20 con artículos demostrativos; 22 con verbos y 23 con tiempos gramaticales y todos los reactivos de la subescala con sustantivos comunes y propios. La única competencia que no estaría representada en el DULM es la de los verbos modales.

A propósito de los reactivos que lograron consenso de los jueces acerca de estar midiendo el conocimiento que debería medir, de acuerdo con el Nodo se puede ver en el segundo renglón que ese es el caso para un reactivo que mide el uso de negativos y otro que valora las preguntas con palabras "*who*" y "*what*.". Dos reactivos logran consenso en cuanto a medir adjetivos comparativos; los nodos de tres reactivos logran el consenso de los jueces respecto al uso de superlativos; cinco de adverbios; ocho del uso de verbos; once de verbos auxiliares y 21 de tiempos gramaticales.

Tabla 5.15. Resultados de las valoraciones de los jueces con respecto a la alineación de los Nodos de la subescala de Gramática con los contenidos del DULM.

| Acuerdos | Sustantivos | Posesivos | Artsdemostrativos | Adjbleous | Adj Comparativos | Adverbios | Superlativos | Pronombres personales | Preposiciones | Conjunciones | Negativos | Who What | Verbos | Tiempos gramaticales | Verbos aux | Verbos modales |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 0 | 0 | 0 | 0 | 2 | 5 | 3 | 0 | 0 | 0 | 1 | 1 | 8 | 21 | 11 | 0 |
| 75 | 2 | 0 | 0 | 0 | 1 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 4 | 0 | 1 | 0 |
| -75 | 32 | 0 | 20 | 3 | 2 | 1 | 0 | 7 | 2 | 1 | 0 | 1 | 10 | 2 | 2 | 0 |
| 0 | 0 | 34 | 2 | 31 | 27 | 28 | 30 | 25 | 32 | 31 | 33 | 32 | 12 | 11 | 20 | 0 |
| Total | 34 | 0 | 20 | 3 | 5 | 6 | 4 | 9 | 2 | 1 | 1 | 2 | 22 | 23 | 14 | 0 |

Enseguida se presenta la Tabla 5.16 que muestra los resultados de la apreciación de las competencias/habilidades de la subescala de Lectura.

Tabla 5.16. Resultados de las valoraciones de los jueces con respecto a la alineación de los Nodos de la subescala de Lectura con los contenidos del DULM.

| A-cuerdos | Info rutinaria | Info no rutinaria | Cartas y notas tema conocido | Instrucciones catálogo | Notas solicitud | Periódico hoteles | Opinión personal | Ensayo tema conocido | Apuntes apoyo lengua materna |
|---|---|---|---|---|---|---|---|---|---|
| 100 | 34 | 32 | 23 | 1 | 24 | 29 | 21 | 26 | 25 |
| 75+ | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 |
| -75 | 0 | 2 | 11 | 29 | 10 | 5 | 13 | 8 | 9 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 34 | 34 | 34 | 34 | 34 | 34 | 34 | 34 | 34 |

Es de notar que todos los reactivos resultan estar relacionados con las competencias del DULM. Lo que es tal vez más importante es que, en esta subescala el número de reactivos que lograron consenso de los jueces acerca de que el Nodo del reactivo es afín a la competencia del DULM es mayor que en las otras dos subescalas. Exceptuando el caso de la competencia de comprensión de instrucciones escritas en catálogos, la cual tiene a un reactivo, las demás competencias agrupan de manera consensuada a un número elevado de reactivos.

**Discusión y conclusiones.**

La indagación de validez de contenido del EXEDII arroja resultados que indican en primer lugar, que una de las mayores dificultades para evaluar el manejo del lenguaje estriba en la delimitación del contenido. Como se dijo en el capítulo 3 de esta tesis el EXEDII es un examen que fue construido con una metodología específica para exámenes orientados a un criterio, razón por la cual está alineado a un currículo. El currículo estaba diseñado desde el enfoque comunicativo, pero supone la existencia de cuatro habilidades, dos productivas y dos receptivas para manejar el inglés como lengua extranjera. Con base en ello, y dado que los recursos para la construcción del examen eran limitados se consideró conveniente que el diseño de los reactivos de la prueba llevara a evaluar los conocimientos de gramática y a medir dos habilidades receptivas: la comprensión auditiva y la de lectura. Ahora bien, si se hubiera seguido la metodología tradicional para validación de contenido se podrían haber comparado los reactivos y los nodos del examen, con el criterio consistente en los contenidos que se enseñan en el curso de nivel intermedio que actualmente se lleva en la Facultad de Idiomas de la UABC. No obstante, se tomó la decisión de trascender esa metodología y sustentar la búsqueda del criterio en una teoría que se apegara más al enfoque comunicativo moderno. Así fue como la búsqueda del estado del arte de la enseñanza y evaluación del inglés como lengua extranjera señaló al modelo de Bachman y Palmer (1996) y a la teoría en la que se sustenta, como una aproximación adecuada para la validación de contenido del EXEDII.

En el apartado 4.5.3 de esta tesis se presenta con amplitud el modelo de Bachman y Palmer, por lo que en este capítulo se asume que el lector está familiarizado con éste. No obstante vale la pena recordar que el modelo propone un marco conceptual para construir o evaluar exámenes de idiomas, en el que las habilidades de los sustentantes en el manejo de un idioma no corresponden a las cuatro habilidades tradicionalmente aceptadas. Más aún, las habilidades del lenguaje evaluadas no ocurren en el vacío, sino que interactúan con el conocimiento del tópico, con el esquema afectivo del sustentante y varían en función de la situación específica. Por ello los autores proponen el concepto del uso de la lengua para encuadrar en éste las características del sustentante, de las tareas y de la situación. De ahí que los exámenes de idiomas deban reflejar esta concepción del uso del lenguaje que incluye las

características esenciales de las situaciones que se supone que enfrentaría el sustentante en el mundo real.

Por lo anterior, el criterio para este estudio es más amplio que el currículo de un curso puesto que se elaboró a partir de las descripciones de las competencias del manejo del inglés como lengua extranjera publicadas en los estándares internacionales, acotadas y adaptadas por las opiniones de expertos en las características de los sustentantes del EXEDII, de las situaciones probables de uso de la lengua meta y de las tareas de esas situaciones. El criterio utilizado constituye la conceptualización del Dominio de Uso de la Lengua Meta.

Es importante hacer la aclaración de que en las descripciones de los estándares internacionales no están diferenciadas las dos habilidades que evalúa el EXEDII (comprensión auditiva y lectura), ni los conocimientos de la estructura de la lengua (gramática) necesarios para ese nivel, aunque en la conceptualización se incluyen habilidades de comunicación que implican la comprensión auditiva dentro de un contexto interactivo en el que se esperaría la producción de discurso; así también la producción de textos escritos implica una respuesta del lector que puede ser evidente o no para el autor del texto, como ocurre en el mundo real. Pero como se dijo arriba es muy difícil separar conceptualmente los componentes de una competencia lingüística cuando se trata de evaluar las habilidades implicadas. Por ello se les pidió a los jueces que para evaluar cada reactivo pensaran en todo momento en situaciones del mundo real que podrían enfrentar sus estudiantes, tomando en cuenta que el EXEDII se enfoca en el segmento receptivo de la competencia descrita en el DULM.

Después de revisar cada uno de los reactivos del EXEDII y evaluar si cumplía con los aspectos mencionados en la Escala de Validación los jueces decidieron que:

- La mayoría de los nodos son considerados relevantes, representativos y pertinentes, por lo que este hallazgo constituye una evidencia de validez de contenido.
- La mayoría de las características de los reactivos son consideradas adecuadas, con excepción de la que se refiere a la capacidad del reactivo de reflejar el

contexto cultural de los sustentantes. Este hallazgo indica que, aunque las características de los reactivos son adecuadas, el no reflejar cabalmente el contexto cultural de los sustentantes puede producir sesgo en las respuestas y por lo tanto, es necesario tener cuidado en las interpretaciones que se hagan de los puntajes y en la generalización de los resultados.

- La subescala de Comprensión auditiva mide las competencias del DULM con excepción de la que establece que el sustentante puede comunicarse con usuarios del idioma que tienen al inglés como lengua nativa. Este hallazgo llama la atención ya que los estímulos de la subescala de comprensión auditiva presentan diálogos entre dos personas que hablan el inglés americano como lengua nativa.
- La mayoría de los estímulos de los reactivos del EXEDII implican el conocimiento y aplicación de la gramática inglesa incluida en el DULM, con excepción del uso del posesivo y los verbos modales.
- Los reactivos de la subescala de Lectura valoran todas las competencias contempladas en el DULM, lo que indica que esta escala es útil para evaluar la comprensión de lectura en inglés al nivel intermedio.
- Los hallazgos de esta investigación constituyen evidencias de que el contenido del EXEDII es adecuado para medir el constructo que se pretende medir, pero es necesario enfatizar que ese constructo representa solamente a una proporción pequeña de las habilidades con que debería contar una persona que requiere comunicarse en inglés, porque no se evalúa la producción del lenguaje.

# 6. VALIDACION DEL CONSTRUCTO DEL EXEDII

El propósito de la investigación que se reporta en esta tesis es el de recabar evidencias de la validez del Examen de Egreso del Idioma Inglés. En este capítulo se presentan el planteamiento, la metodología y los resultados obtenidos en la indagación de la validez de constructo de ese examen.

La estrategia de indagación que se describe en este capítulo parte de dos preguntas de investigación planteadas en el apartado 1.4. La primera pregunta se retoma aquí para mayor comprensión de la estrategia de respuesta:

> ¿En qué grado el EXEDII constituye una escala de reactivos que miden conjuntamente un constructo o dimensión y de qué manera estos datos aportan evidencias de validez de constructo?

El primer punto que se requiere investigar es si el EXEDII es un instrumento que mide un solo constructo. Se recordará que el constructo del examen está definido en términos generales como el manejo del inglés como lengua extranjera, al nivel intermedio en los estudiantes que egresan de la UABC. Al mismo tiempo, la definición del constructo constituye el supuesto fundamental de todo el trabajo de tesis.

Desde el punto de vista de la evaluación educativa un examen es un instrumento de medición que evalúa los conocimientos, habilidades o competencias de sus sustentantes en un tema particular. Eso implica que es posible conceptualizar y definir operacionalmente el constructo que supuestamente mide el instrumento y que mediante técnicas adecuadas es posible evaluarlo. La validez de las interpretaciones que se hagan acerca de los puntajes obtenidos en un examen así concebido depende directamente del grado en el que se asegure esa congruencia.

## 6.1. Planteamiento del problema.

La definición del constructo del EXEDII mencionada arriba es un planteamiento abstracto que responde a los aspectos teóricos de los que se deriva el diseño del instrumento. Pero en tanto que es un instrumento de medición de un constructo, se puede considerar como un conjunto de habilidades y conocimientos, funcionando en combinación para responder a situaciones particulares en el mundo real.

Desde el punto de vista metodológico debería ser posible demostrar primeramente que todos los reactivos del EXEDII miden una misma dimensión.

La segunda pregunta de investigación dice:

> ¿En qué grado la estructura factorial del EXEDII es congruente con el diseño de la prueba y de qué manera estos datos aportan evidencias de validez de constructo?

Por lo tanto, se requiere investigar es si las relaciones entre un número de variables observadas pueden ser explicadas en términos de un número menor de variables no observadas.

Por lo anterior las evidencias de validez de constructo del EXEDII se recabaron a través de una combinación de procedimientos estadísticos: en primer lugar se investigó la dimensionalidad del instrumento y en segundo lugar se indagaron los patrones de convergencia de diferentes medidas representadas por las respuestas a los reactivos del EXEDII.

La estrategia general estuvo planeada desde la perspectiva de la Teoría de Respuesta al Ítem (TRI) porque ésta constituye un marco útil para resolver una variedad de problemas de medición, algunos de los cuales se explican más adelante. Por lo tanto la dimensionalidad del instrumento se valoró utilizando el Modelamiento Rasch y la estructura factorial del examen mediante la determinación de los patrones de convergencia de las diferentes medidas de los

reactivos del EXEDII. En el siguiente apartado se expone el proceso de la indagación de la dimensionalidad del EXEDII.

## 6.2. Indagación de la Dimensionalidad.

El modelo Rasch es el análisis más elemental de la Teoría de Respuesta al Reactivo. Tiene dos supuestos fundamentales: la unidimensionalidad y la independencia local. El primero especifica que un conjunto dado de reactivos mide una sola habilidad. Pero ello no significa que el desempeño del sustentante de una prueba se deba exclusivamente a un solo proceso cognoscitivo como una habilidad aislada. Particularmente en el tipo de constructo que mide el EXEDII (dominio de una lengua extranjera) no pueden suponerse habilidades que actúan de manera aislada; más bien se supone que un conjunto de ellas determinan la respuesta a cada tarea del dominio o reactivo de la prueba. Pero al demostrar que los reactivos del examen funcionan en conjunto se obtiene evidencia de una habilidad general y las medidas obtenidas se comportan como componentes que definen el constructo de interés (D´Agostino, citado en González Montesinos, 2008). El supuesto de la independencia local especifica que cuando las habilidades que influyen sobre el desempeño en la prueba se mantienen constantes las respuestas de los sustentantes son estadísticamente independientes, en cualquier par de reactivos determinados.

El Modelamiento Rasch se realiza sobre los datos observados utilizando un método denominado Máxima Verosimilitud que proporciona las estimaciones de los parámetros que con mayor probabilidad hubieran producido los patrones de respuesta observados en los datos (González Montesinos, 2008). El análisis produce una curva característica del reactivo (CCR) que se obtiene combinando los parámetros de habilidad y dificultad a través de su diferencia (Tristán, 2002). Siguiendo la ecuación característica se establece que en el encuentro del sustentante con el reactivo, para los sustentantes con habilidad mayor que la dificultad del reactivo la diferencia es positiva y tienen una probabilidad mayor a 0.50 de acertar, mientras que cuando la dificultad del reactivo es mayor que la habilidad del sustentante la diferencia es

negativa y la probabilidad de una respuesta correcta es menor a 0.50 (González Montesinos, 2008).

### 6.2.1. Modelamiento Rasch

A continuación se explican los detalles metodológicos de la estrategia de validación de constructo. Por razones de espacio se presentan en este capítulo las tablas que resumen los resultados, pero en el Anexo 5.1 se pueden consultar los datos completos.

#### 6.2.1.1. Método.

**Participantes.**

Los datos analizados corresponden a los resultados obtenidos en el EXEDII por los sustentantes de las cohortes de 2005 y 2006. Para el análisis se prepararon de antemano dos archivos: el archivo de datos, capturados en un solo archivo con extensión txt y el archivo de control que consiste en una serie de instrucciones en las que se especifican las características del análisis que se desea realizar (Ver Anexo 5.1). La base de datos se construyó en el programa de cómputo *UltraEdit-32* (*IDM Computer Solutions, Inc.*), la cual está constituida por las respuestas a los cien reactivos del EXEDII (ordenados por columnas) obtenidas de los 2260 sustentantes (ordenados en los renglones).

**Procedimiento.**

El Análisis de Rasch de las respuestas de los sustentantes del EXEDII se realizó mediante el programa de cómputo Winsteps (Linacre, 2003, 2006) el cual efectúa primeramente un estimado central para cada persona y una calibración del reactivo, así como una calibración respuesta-estructura y las medidas son reportadas en lógitos. Un lógito es una unidad de medida especial que se obtiene multiplicando los momios de respuesta a cada reactivo por el logaritmo natural Ln (Tristán, 2002). La utilidad de utilizar una escala de esta naturaleza es que permite expresar el parámetro de la dificultad del reactivo en una escala uniforme. Igualmente, la escala

en lógitos se aplica para caracterizar los grados de habilidad de los sustentantes (Tristán 2002, p. 11-15).

Una vez que el programa realiza el estimado para personas y reactivos procede a realizar una serie de iteraciones con el algoritmo UCON para alcanzar una convergencia lo más aproximada posible con el patrón de datos observados. Los estadígrafos de ajuste permiten detectar las "anomalías" o datos acerca de los reactivos que no se ajustan al modelo lo que permite su revisión para ser modificados o eliminados.

Efectuar un de Análisis Rasch para los propósitos de este estudio es pertinente porque proporciona: a) la Medida que es la calibración de la dificultad del reactivo expresada en lógitos; b) el error estándar de la media en lógitos; c) los valores del ajuste interno (*INFIT*) y externo (*OUTFIT*); d) la correlación punto-biserial entre cada reactivo calificado dicotómicamente y la puntuación total observada para el reactivo; e) la discriminación de cada reactivo para distinguir entre sustentantes de alta y baja habilidad.

El archivo de salida del Modelamiento Rasch consiste de una variedad de gráficos y tablas y los estadígrafos de ajuste se reportan como residuales de la media cuadrática (MNSQ), mismos que tienen una distribución aproximada de Chi cuadrada, y también se reportan como t estandarizada (ZSTD) (Linacre, 2006). Con la ayuda de los criterios de Bondad de Ajuste recomendados en la literatura y en el propio manual del programa Winsteps se detectan comportamientos de respuesta no esperados o anomalías de ajuste interno, mientras que con otros criterios se detectaron comportamientos extremos de ajuste externo.

#### 6.2.1.2 Resultados.

En el Anexo 5.1 se presenta el archivo de salida en extenso del Análisis de Rasch realizado. Ese anexo contiene seis apartados en los que se incluyen: A) el archivo de control utilizado para correr el análisis; B) la tabla de convergencia; C) las estadísticas sumarias del análisis; D) la tabla de calibración de los reactivos con los valores de los estadígrafos de ajuste

al modelo; E) la tabla de las medidas de los sustentantes con los valores de los estadígrafos de ajuste al modelo; F) gráficas. En este capítulo se ofrece un resumen de los resultados.

De acuerdo con la tabla de convergencia obtenida como parte de los archivos de salida que proporciona el programa se tiene que:

Cuatro iteraciones fueron necesarias antes de lograr convergencia.

El puntaje promedio es de 57.0 y su desviación estándar es de 20.3.

El menú principal del archivo de salida permite seleccionar las tablas que se requieren para interpretar los resultados del análisis de acuerdo con el propósito del estudio. Los criterios para la interpretación fueron consultados en el artículo de González Montesinos (2008) y directamente del manual de Winsteps (1991-2006). A continuación se presentan los resultados totales para la prueba.

**Estadísticas descriptivas sumarias**.

- La dificultad promedio se ubica en .00 lógitos (desviación estándar de .90).
- La medida de la habilidad promedio de los sustentantes es de .45 lógitos (desviación estándar 1.17).
- El error promedio de los reactivos es de .05, que es un valor bajo.
- El error promedio es de los sustentantes es de .25, que también es un valor bajo considerando que se refiere a los sustentantes.
- El estadígrafo de ajuste interno (*INFIT*) es sensible a comportamientos inesperados con respecto al modelo Rasch, que afectan a las respuestas a aquellos reactivos cercanos al nivel de habilidad del sustentante y se obtiene un valor promedio de 1.
- Los indicadores de ajuste al modelo son *MNSQ*, que es la media cuadrática y *ZSTD* que es el mismo estadígrafo, pero estandarizado. Para interpretar los valores obtenidos en la literatura se reportan valores que funcionan como criterios para decidir cuándo un reactivo no se ajusta al modelo. Por ejemplo Bond y Fox (citado en González Montesinos, 2008) dicen que siendo 1 el valor

esperado, el hallazgo de valores superiores a 1 indican falta de juste al modelo. Un valor de 1.30 indica 30% más de variación entre el modelo y los valores obtenidos. Valores menores a 1 indican menor variación que la esperada, por lo que un valor de .80 indica una variación 20% menor a la esperada. No obstante, el apego a estos valores criterio puede decidirse dependiendo de las características de la investigación. En este estudio se toman como valores criterio un rango entre -1.3 y 1.0 para la media cuadrática (*MNSQ*) y -2 a +2 para el estadígrafo estandarizado (ZSTD).

- El estadígrafo de ajuste externo (*OUTFIT*) es sensible a comportamientos inesperados que afectan las respuestas a los reactivos lejanos al nivel de habilidad del sustentante. Los valores criterio para *MNSQ* y *ZSTD* son los mismos que para el ajuste interno, habiéndose obtenido un valor promedio de 1.
- El estadígrafo que informa sobre el error total es la raíz del error cuadrático promedio (*RMSE*) y proporciona información sobre la cantidad de variación aleatoria en una muestra. El error cuadrático promedio real (*Real RMSEA*) se calcula considerando los desajustes de los datos que se desvían del modelo y se interpreta como el "peor caso de confiabilidad" (límite inferior). Se obtuvo un valor promedio de .05
- El error del modelo (*Model RMSEA*) es también la raíz del error cuadrático promedio se reporta como el límite superior de los estimados de confiabilidad, o el mejor caso. En este caso, los dos valores coinciden en el caso de los reactivos (.05), así como para los sustentantes (.26).
- El estimado de la desviación estándar "verdadera" es la que se obtiene después de sustraer al valor obtenido el sesgo derivado del error de medición. En este caso los valores esperados coinciden con los obtenidos (.90).
- El error estándar de la media de las personas (.09) y de los reactivos (.02) presentan valores bajos.
- Para el índice de confiabilidad (Alfa de Cronbach) se obtiene un valor de .95 considerado alto. Este indicador informa sobre la estabilidad de la medición.

- El coeficiente de correlación de Pearson entre los puntajes crudos y las medidas en lógitos, incluyendo los puntajes extremos. El criterio de ajuste indica que deben estar cerca de -1. Se obtuvo un valor de -1.0.

**Estadísticas de los reactivos**.

En el archivo de salida para los reactivos se puede observar la calibración de cada uno de ellos y sus estadígrafos de ajuste al modelo. Los hallazgos más importantes se presentan a continuación, pero pueden consultarse en extenso en el Anexo 6.

- La calibración de los reactivos se observa en un rango de -2.01 a +1.80 lógitos.
- El error de los reactivos (model error) es de .05 para 88 reactivos, mientras que nueve muestran error de .06 y tres de .07.
- Los estadígrafos de ajuste interno y externo muestran que la mayor parte de los reactivos presentan valores dentro de los rangos del criterio de bondad de ajuste. No obstante, 13 reactivos presentan valores fuera de esos rangos. En la Tabla 6.1 se muestran estos reactivos y los valores obtenidos.

Tabla 6.1 Reactivos que presentan anomalías en los indicadores de ajuste al modelo Rasch, con valores fuera de los rangos esperados.

| Reactivos | Ajuste interno | | Ajuste externo | |
|---|---|---|---|---|
| | MNSQ 1 a 1.3 | ZSTD -1 a +2 | MNSQ 1 a 1.3 | ZSTD -1 a +2 |
| 16 | 1.30 | 2.1 | 1.75 | 2.2 |
| 62 | 1.31 | 2.1 | 1.45 | 2.4 |
| 76 | 1.34 | 2.6 | 1.41 | 2.3 |
| 52 | 1.34 | 2.5 | 1.45 | 2.5 |
| 59 | | -2.2 | | -2.2 |
| 30 | | -2.1 | | |
| 12 | | | 1.40 | |
| 34 | | | 1.33 | |
| 9 | | | 2.42 | |
| 41 | | | 1.38 | |
| 70 | | | 1.32 | |
| 79 | | | 1.39 | |
| 86 | | | 1.43 | |

**Estadísticas de los sustentantes.**

Los resultados observados para los sustentantes se pueden consultar en el Anexo 6. Por razones de espacio en el anexo solamente se incluyen los 200 con medida más alta y los 200 con medida más baja. En el archivo de salida se observa que el error estándar de cada medida (*model error)* presenta valores entre .30 y 1.01. Los valores más altos están asociados a las medidas de 4 lógitos y tienden a disminuir desde .72 hasta .24 para los sustentantes con medida de -1 lógito. Vuelven a aumentar paulatinamente hasta .30 con el sustentante de menor medida de habilidad (-2.09) lógitos, por lo que se observa que los errores más elevados se asocian con las medidas extremas de habilidad.

La medida de los sustentantes se observa en un rango de -2.09 a +4.9 lógitos. La distribución de la habilidad se presenta enseguida, en el mapa que relaciona la alineación de los reactivos con la habilidad de los sustentantes, organizados por grupos de habilidad (Figura 6.1).

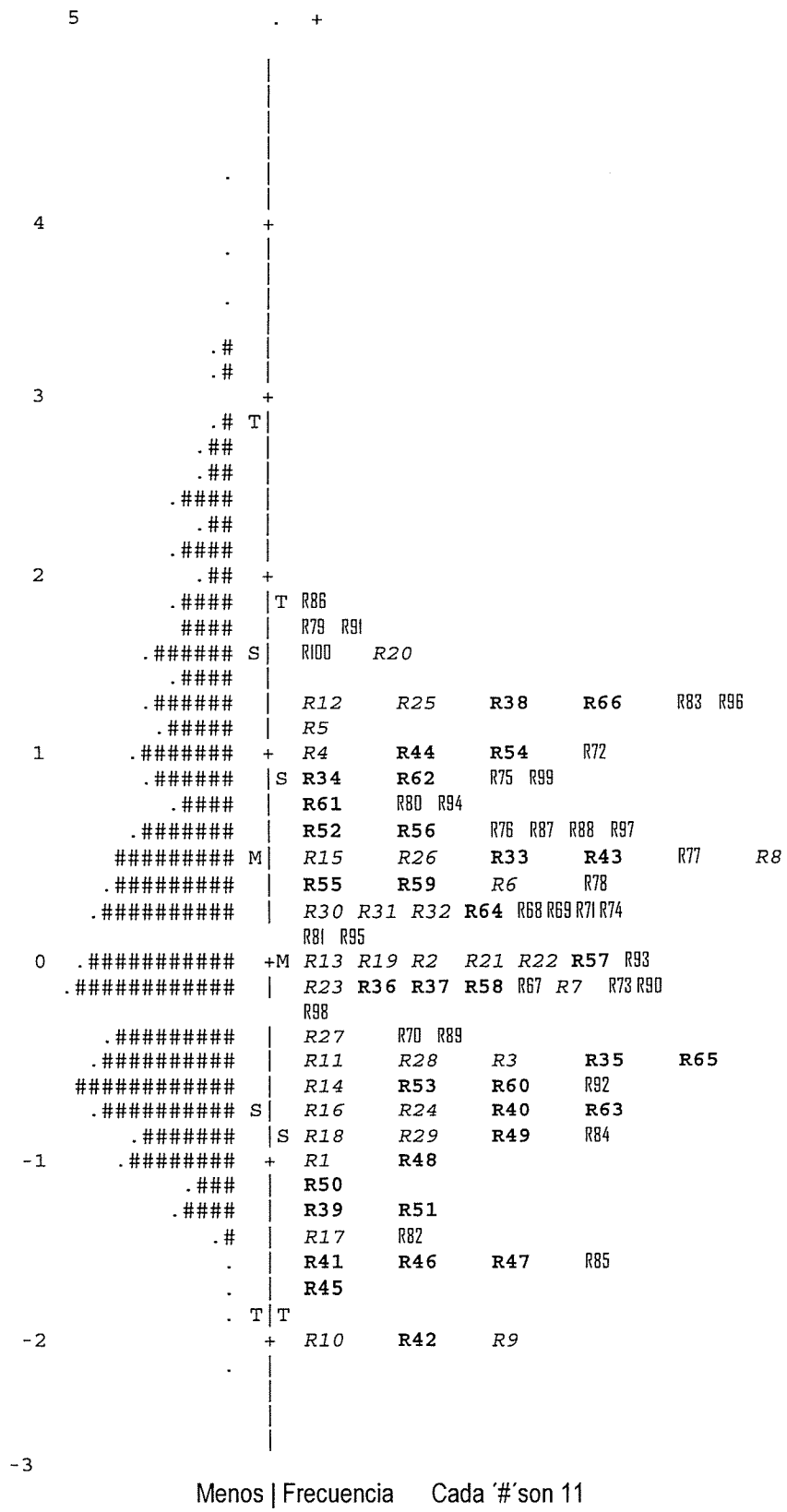



Figura 6.1. Mapa de la relación dificultad de los reactivos con la habilidad de los sustentantes.

El mapa constituye un indicador del ajuste al modelo de los datos observados; pueden notarse visualmente características tales como que los reactivos se distribuyen de acuerdo con la curva normal, con la mayor parte de ellos con una calibración de 0 lógitos, aumentando paulatinamente hacia +1 y disminuyendo hacia -1 lógitos, en su mayoría. Para distinguir los reactivos de acuerdo con la subescala a la que corresponden, los de Comprensión auditiva se han escrito en itálicas, los de Gramática aparecen en negritas y los de Lectura aparecen en un tipo de letra diferente. En el mapa se observa que:

- La mayor parte de los reactivos se agrupan entre -1 y +2 lógitos de habilidad calibrada aunque se observan sustentantes con habilidad de 3, 4 y 5 lógitos.
- La distribución de los reactivos en el mapa permite ubicar la dificultad de los reactivos en relación con la medida de habilidad de los sustentantes.
- Los reactivos de las tres subescalas del EXEDII se ubican en el rango de dificultad de -1 a +2 lógitos.
- En -2 lógitos y -1.99 se encuentran los tres reactivos más fáciles: dos de Comprensión auditiva y uno de Gramática. En el otro extremo, +1.80 se encuentran los cuatro reactivos más difíciles: tres de Lectura y uno de Comprensión auditiva.
- Entre 0 y +1 lógitos hay 15 reactivos de Comprensión auditiva, 15 de Gramática y 23 de Lectura.
- Entre 0 y -1 lógitos hay 17 reactivos de Comprensión auditiva, 18 de Gramática y 11 de Lectura.
- Dos reactivos tienen calibración de -2 lógitos.

Con respecto al ajuste al modelo Rasch de los patrones de respuesta analizados el estadígrafo de ajuste interno (INFIT) muestra que la mayor parte de los sustentantes obtienen valores dentro del rango esperado para la media cuadrática (MNSQ). No obstante, algunos sustentantes presentan valores superiores a 1.30. En el estadígrafo de ajuste interno estandarizado (ZSTD) ocurre algo similar, ya que la mayoría de los valores caen dentro del rango esperado, con excepción de algunos que presentan valores mayores a 2 o menores a -2.

Los estadígrafos de ajuste externo (*OUTFIT*) tanto la media cuadrática, como la estandarizada presentan la mayor parte de los reactivos dentro de los valores esperados. Sin embargo, algunos sustentantes obtienen valores fuera del rango esperado. En la Tabla 6.2 se presentan las medidas de la habilidad de los sustentantes y los valores correspondientes para los estadígrafos, en los casos en los que éstos rebasan los criterios de bondad de ajuste. Estos últimos se muestran en negritas.

Así, puede notarse que de los 2260 sustentantes que conforman la muestra, 225 obtienen resultados con valores fuera de la expectativa del modelo, 37 presentan desajuste en los cuatro estadígrafos de ajuste, por lo que puede decirse que no pertenecen a la escala. Estos sustentantes corresponden a medidas de habilidad entre +1 lógito (1 sustentante); 0 lógitos (30); y -1 lógito (6 sustentantes). Se marcan con un asterisco (*) en la tabla.

Por otra parte, 84 sustentantes presentan valores fuera de lo esperado solamente en la media cuadrática de ajuste interno; 75 en el estadígrafo estandarizado, mientras que 202 tienen valores fuera del rango esperado para la media cuadrática de ajuste externo y 88 para la media cuadrática estandarizada de ajuste externo.

Tabla 6.2. Resultados de los sustentantes que no se ajustan a los valores esperados en el modelo Rasch

| Medida en lógitos | Media cuadrática MNSQ 1 a 1.3 | Media cuadrática estandarizada ZSTD -2 a +2 | Media cuadrática MNSQ 1 a 1.3 | Media cuadrática estandarizada ZSTD -2 a +2 | | Medida en lógitos | Media cuadrática MNSQ 1 a 1.3 | Media cuadrática estandarizada ZSTD -2 a +2 | Media cuadrática MNSQ 1 a 1.3 | Media cuadrática estandarizada ZSTD -2 a +2 |
|---|---|---|---|---|---|---|---|---|---|---|
| 4.24 | 1.04 | .1 | **1.98** | .7 | | -.04 | **.86** | **2.1** | **.82** | 1.8 |
| 3.27 | 1.03 | .1 | **2.61** | 1.7 | | -.04 | 1.16 | **2.1** | 1.20 | 1.8 |
| 3.07 | **.96** | -.1 | **2.19** | 1.5 | | -.08 | **.84** | **2.3** | **.81** | 1.9 |
| 3.07 | 1.04 | .1 | **1.56** | .8 | | -.08 | **.82** | **2.7** | **.81** | 1.9 |
| 3.07 | 1.09 | .3 | **1.54** | .8 | | -.08 * | **.84** | **2.4** | **.78** | **2.3** |
| 3.07 | .99 | .0 | **2.18** | 1.5 | | -.08 | **.87** | 2.0 | **.81** | 2.0 |
| 3.07 | 1.09 | .3 | **3.23** | **2.5** | | -.08 | 1.23 | **3.0** | **1.34** | **2.9** |
| 3.07 | 1.02 | .1 | **2.26** | 1.6 | | -.08 * | **.83** | **2.5** | **.79** | **2.1** |
| 3.07 | 1.09 | .3 | **3.23** | **2.5** | | -.08 * | **.82** | **2.7** | **.76** | **2.5** |
| 2.90 | 1.04 | .1 | **1.98** | 1.4 | | -.08 * | **.80** | **3.1** | **.75** | **2.6** |
| 2.90 | **.99** | .0 | **2.03** | 1.5 | | -.13 * | **.84** | **2.3** | **.80** | **2.1** |
| 2.90 | 1.00 | .0 | **1.58** | .9 | | -.13 * | **.85** | **2.2** | **.80** | **2.1** |
| 2.75 | **.95** | -.2 | **1.73** | 1.2 | | -.13 | **.85** | **2.2** | **.82** | 1.9 |
| 2.75 | 1.05 | .2 | **1.88** | 1.4 | | -.13 * | **.81** | **2.8** | **.76** | **2.5** |
| 2.75 | **.93** | -.3 | **1.60** | 1.0 | | -.18 * | **.80** | **2.9** | **.76** | **2.5** |
| 2.75 | 1.06 | .2 | **1.96** | 1.5 | | -.23 | 1.27 | **3.4** | **1.40** | **3.3** |
| 2.75 | 1.07 | .3 | **1.39** | .7 | | -.23 | **.85** | **2.2** | **.81** | 1.9 |
| 2.75 | 1.05 | .2 | **1.49** | .9 | | -.23 | **.83** | **2.4** | **.81** | 1.9 |
| 2.61 | **.96** | -.2 | **1.55** | 1.0 | | -.32 | **.84** | **2.3** | **.85** | 1.4 |
| 2.61 | **.89** | -.5 | **1.44** | .8 | | -.32 | **.83** | **2.3** | **.81** | 1.8 |
| 2.61 | 1.01 | .0 | **1.35** | .7 | | -.37 * | **.73** | **3.9** | **.67** | **3.4** |
| 2.61 | **.98** | -.1 | **1.62** | 1.1 | | -.37 | **.83** | **2.3** | **.80** | 2.0 |
| 2.49 | 1.15 | .6 | **2.34** | **2.3** | | -.37 | 1.18 | **2.2** | 1.26 | 2.1 |
| 2.49 | 1.04 | .2 | **2.06** | 1.9 | | -.42 | 1.28 | **3.2** | **1.41** | **3.2** |
| 2.49 | 1.01 | .0 | **1.57** | 1.1 | | -.42 | 1.22 | **2.6** | 1.26 | **2.1** |
| 2.49 | **.91** | -.4 | **1.38** | .8 | | -.47 | 1.24 | **2.8** | **1.31** | **2.4** |
| 2.38 | 1.05 | .2 | **1.80** | 1.6 | | -.47 | **.83** | **2.2** | **.87** | 1.2 |
| 2.38 | **.92** | -.4 | **1.38** | .9 | | -.52 | 1.25 | **2.9** | **1.34** | **2.6** |
| 2.38 | 1.08 | .4 | **1.44** | 1.0 | | -.57 * | **.80** | **2.6** | **.74** | **2.3** |
| 2.38 | 1.06 | .3 | **2.01** | 2.0 | | -.57 * | 1.23 | **2.5** | **1.41** | **2.9** |
| 2.27 | **.98** | -.1 | **1.31** | .8 | | -.57 | 1.24 | **2.7** | 1.30 | **2.2** |
| 2.27 | **.98** | -.1 | **1.31** | .8 | | -.57 | 1.19 | **2.2** | 1.22 | 1.7 |
| 2.27 | **.98** | -.1 | **1.38** | .9 | | -.57 * | **1.34** | **3.7** | **1.41** | **2.9** |
| 2.27 | **.91** | -.5 | **1.65** | 1.4 | | -.62 | 1.24 | **2.6** | **1.34** | **2.4** |
| 2.17 | 1.03 | .1 | **1.41** | 1.0 | | -.62 | 1.16 | 1.8 | **1.35** | **2.5** |
| 2.17 | 1.06 | .3 | **1.50** | 1.2 | | -.62 | 1.21 | **2.3** | **1.42** | **2.9** |
| 2.17 | 1.03 | .2 | **1.50** | 1.2 | | -.67 | 1.26 | **2.7** | **1.38** | **2.6** |
| 2.08 | 1.04 | .2 | **1.44** | 1.1 | | -.67 | 1.26 | **2.7** | **1.38** | **2.6** |
| 1.99 | 1.03 | .2 | **1.55** | 1.5 | | -.67 | 1.30 | **3.2** | **1.53** | **3.4** |
| 1.99 | 1.03 | .2 | **1.55** | 1.5 | | -.72 | 1.23 | **2.4** | **1.36** | **2.4** |
| 1.99 | 1.07 | .4 | **1.69** | 1.8 | | -.72 | 1.21 | **2.2** | **1.31** | **2.1** |
| 1.99 | 1.07 | .4 | **1.43** | 1.2 | | -.72 | 1.20 | **2.1** | **1.31** | 2.0 |
| 1.99 | 1.07 | .4 | **1.43** | 1.2 | | -.72 | 1.25 | **2.6** | **1.35** | **2.3** |
| 1.83 | **.96** | -.3 | **1.57** | 1.7 | | -.72 | 1.11 | 1.2 | **1.35** | **2.3** |
| 1.75 | 1.01 | .1 | **1.42** | 1.3 | | -.72 | 1.11 | 1.2 | **1.35** | **2.3** |
| 1.68 | 1.08 | .5 | **1.46** | 1.5 | | -.72 | 1.22 | **2.3** | **1.34** | **2.2** |
| 1.68 | 1.06 | .4 | **1.32** | 1.1 | | -.72 | 1.24 | **2.5** | **1.36** | **2.3** |
| 1.68 | 1.09 | .6 | **1.31** | 1.1 | | -.77 | 1.30 | **3.0** | **1.50** | **3.0** |
| 1.61 | 1.08 | .6 | **1.31** | 1.1 | | -.77 | 1.26 | **2.6** | **1.35** | **2.2** |
| 1.54 | 1.11 | .8 | **1.32** | 1.2 | | -.77 | 1.24 | **2.4** | **1.34** | **2.2** |
| 1.54 | 1.18 | 1.3 | **1.41** | 1.5 | | -.82 | 1.28 | **2.7** | **1.32** | 2.0 |
| 1.54 | 1.15 | 1.1 | **1.39** | 1.4 | | -.88 | 1.20 | 1.9 | **1.39** | **2.3** |
| 1.54 | 1.07 | .6 | 1.30 | 1.1 | | -.88 | 1.22 | **2.1** | **1.32** | 1.9 |
| 1.54 | 1.18 | 1.3 | **1.56** | 2.0 | | -.88 | 1.22 | **2.1** | 1.26 | 1.6 |
| 1.54 | 1.10 | .8 | **1.47** | 1.7 | | -.88 | 1.21 | 2.0 | 1.27 | 1.6 |
| 1.47 | 1.00 | .0 | **1.35** | 1.4 | | -.88 | 1.20 | 2.0 | 1.27 | 1.6 |
| 1.47 | 1.10 | .8 | **1.31** | 1.2 | | -.88 | 1.20 | 2.0 | **1.40** | **2.3** |
| 1.47 | 1.02 | .2 | 1.30 | 1.2 | | -.88 | 1.17 | 1.6 | **1.38** | **2.2** |
| 1.41 | 1.02 | .2 | **1.36** | 1.5 | | -.88 | 1.23 | **2.2** | **1.37** | **2.2** |
| 1.22 | 1.14 | 1.2 | **1.51** | **2.2** | | -.93 | 1.29 | **2.6** | **1.48** | **2.6** |
| 1.22 | 1.06 | .5 | **1.41** | 1.8 | | -.93 * | **1.31** | **2.8** | **1.60** | **3.2** |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1.05 | 1.15 | 1.5 | 1.33 | 1.7 | | -.93 | 1.20 | 1.9 | 1.38 | 2.1 |
| 1.05 * | .78 | 2.4 | .66 | 2.2 | | -.93 | 1.21 | 2.0 | 1.38 | 2.1 |
| 1.05 | .82 | 1.9 | .74 | 1.6 | | -.93 | 1.18 | 1.7 | 1.36 | 2.1 |
| .94 | 1.16 | 1.6 | 1.44 | 2.4 | | -.99 * | 1.35 | 3.1 | 1.47 | 2.5 |
| .79 * | .80 | 2.5 | .72 | 2.1 | | -.99 * | 1.32 | 2.8 | 1.51 | 2.7 |
| .73 | .84 | 2.0 | .76 | 1.8 | | -.99 | 1.19 | 1.7 | 1.37 | 2.0 |
| .68 | .84 | 2.0 | .77 | 1.8 | | -.99 | 1.20 | 1.8 | 1.30 | 1.7 |
| .68 | .82 | 2.4 | .76 | 1.9 | | -.99 | 1.27 | 2.4 | 1.34 | 1.9 |
| .63 * | 1.32 | 3.6 | 1.59 | 3.7 | | -.99 * | 1.33 | 2.9 | 1.59 | 3.0 |
| .63 | .83 | 2.3 | .77 | 1.9 | | -1.04 | 1.27 | 2.4 | 1.38 | 2.0 |
| .58 * | .78 | 3.1 | .71 | 2.5 | | -1.04 | 1.11 | 1.0 | 1.37 | 2.0 |
| .58 * | .80 | 2.7 | .73 | 2.3 | | -1.04 | 1.11 | 1.0 | 1.37 | 2.0 |
| .58 * | .83 | 2.3 | .76 | 2.1 | | -1.04 | 1.11 | 1.0 | 1.31 | 1.6 |
| .54 | 1.16 | 2.0 | 1.25 | 1.8 | | -1.04 | 1.22 | 2.0 | 1.30 | 1.6 |
| .54 | 1.19 | 2.4 | 1.16 | 1.2 | | -1.04 | 1.27 | 2.4 | 1.41 | 2.1 |
| .54 | 1.18 | 2.2 | 1.34 | 2.4 | | -1.04 | 1.16 | 1.5 | 1.30 | 1.6 |
| .44 | .86 | 2.0 | .79 | 1.9 | | -1.10 | 1.17 | 1.5 | 1.47 | 2.3 |
| .44 * | .84 | 2.3 | .77 | 2.1 | | -1.10 * | 1.33 | 2.8 | 1.54 | 2.6 |
| .44 * | .77 | 3.4 | .71 | 2.7 | | -1.10 | 1.20 | 1.7 | 1.53 | 2.6 |
| .39 | .85 | 2.1 | .86 | 1.3 | | -1.10 * | 1.33 | 2.8 | 1.51 | 2.5 |
| .39 | .84 | 2.3 | .78 | 2.0 | | -1.10 | 1.08 | .7 | 1.36 | 1.8 |
| .39 | 1.16 | 2.1 | 1.15 | 1.2 | | -1.16 | 1.22 | 1.8 | 1.30 | 1.5 |
| .34 | 1.24 | 3.1 | 1.30 | 2.4 | | -1.16 | 1.15 | 1.3 | 1.36 | 1.8 |
| .34 * | .84 | 2.3 | .77 | 2.2 | | -1.22 | 1.13 | 1.1 | 1.32 | 1.5 |
| .34 * | .84 | 2.4 | .78 | 2.1 | | -1.22 * | 1.35 | 2.7 | 1.62 | 2.7 |
| .29 | 1.20 | 2.6 | 1.18 | 1.5 | | -1.22 * | 1.33 | 2.6 | 1.73 | 3.1 |
| .29 | .85 | 2.2 | .81 | 1.8 | | -1.22 | 1.26 | 2.1 | 1.52 | 2.3 |
| .29 * | .82 | 2.7 | .77 | 2.3 | | -1.22 | 1.20 | 1.6 | 1.38 | 1.8 |
| .25 | 1.18 | 2.4 | 1.16 | 1.4 | | -1.22 | 1.15 | 1.2 | 1.48 | 2.2 |
| .25 * | .79 | 3.2 | .76 | 2.4 | | -1.22 | 1.15 | 1.2 | 1.34 | 1.6 |
| .20 * | .84 | 2.3 | .79 | 2.1 | | -1.22 | 1.14 | 1.2 | 1.36 | 1.7 |
| .20 | .86 | 2.0 | .83 | 1.7 | | -1.22 | 1.25 | 2.0 | 1.41 | 1.9 |
| .20 | 1.23 | 3.0 | 1.29 | 2.5 | | -1.28 | 1.19 | 1.5 | 1.60 | 2.5 |
| .20 | .86 | 2.0 | .82 | 1.7 | | -1.28 | 1.11 | .9 | 1.35 | 1.6 |
| .20 | .86 | 2.0 | .82 | 1.7 | | -1.34 * | 1.31 | 2.3 | 1.64 | 2.6 |
| .15 | .85 | 2.3 | .80 | 2.0 | | -1.34 | 1.14 | 1.1 | 1.42 | 1.8 |
| .11 | .86 | 2.1 | .81 | 2.0 | | -1.34 | 1.14 | 1.1 | 1.42 | 1.8 |
| .11 * | .83 | 2.6 | .80 | 2.1 | | -1.40 | 1.22 | 1.6 | 1.50 | 2.0 |
| .11 | 1.20 | 2.7 | 1.28 | 2.4 | | -1.40 | 1.24 | 1.8 | 1.56 | 2.2 |
| .11 | .85 | 2.2 | .84 | 1.7 | | -1.40 | 1.18 | 1.3 | 1.38 | 1.6 |
| .06 | .85 | 2.2 | .82 | 1.9 | | -1.40 | 1.18 | 1.3 | 1.38 | 1.6 |
| .06 * | .81 | 2.9 | .79 | 2.2 | | -1.47 * | 1.35 | 2.4 | 1.58 | 2.2 |
| .06 | .86 | 2.1 | .81 | 2.0 | | -1.47 | 1.24 | 1.7 | 1.51 | 2.0 |
| .06 | .84 | 2.4 | .80 | 2.0 | | -1.47 | 1.30 | 2.1 | 1.44 | 1.7 |
| .06 | 1.18 | 2.4 | 1.27 | 2.4 | | -1.54 | 1.22 | 1.5 | 1.38 | 1.4 |
| .01 | .86 | 2.1 | .83 | 1.7 | | -1.54 | 1.22 | 1.5 | 1.38 | 1.4 |
| -.04 | 1.16 | 2.1 | 1.23 | 2.0 | | -1.61 | 1.12 | .8 | 1.35 | 1.3 |
| -.04 | .86 | 2.1 | .83 | 1.7 | | -1.61 | 1.20 | 1.3 | 1.61 | 2.1 |
| -.04 | .85 | 2.2 | .83 | 1.7 | | -1.61 | 1.21 | 1.4 | 1.63 | 2.1 |
| -.04 | .83 | 2.5 | .81 | 2.0 | | -1.76 | 1.25 | 1.5 | 1.76 | 2.3 |
| -.04 * | .81 | 2.9 | .78 | 2.3 | | -1.76 | 1.26 | 1.5 | 1.65 | 2.0 |

N = 225 sustentantes

Otro indicador de la forma en la que trabajó el examen con esta muestra en particular es la que proporciona el resultado que se observa en la Tabla 6.3, en la cual se presenta el resumen del comportamiento de las categorías de respuesta.

Hay dos categorías de respuesta reportadas en la primera columna (0 y 1) con los porcentajes empíricamente observados para cada una de ellas, así como con los porcentajes esperados de acuerdo con el modelo (3ª y 4ª columnas). Se observa que los valores coinciden.

La Tabla 6.3 presenta el comportamiento de las categorías de respuesta.

| 2260 personas | | | | 100 reactivos | | | |
|---|---|---|---|---|---|---|---|
| | | | | Ajuste interno | Ajuste externo | Coherencia | |
| Categoría de respuesta | % | % observado | Expectativa muestra | MNSQ | MNSQ | M->C | C->M |
| 0 | 43 | -.43 | -.43 | .99 | .97 | 69% | 67% |
| 1 | 57 | 1.11 | 1.11 | 1.01 | 1.05 | 76% | 78% |

La tabla presenta también los promedios de ajuste interno y externo de cada categoría. Se observan valores dentro de la expectativa del modelo (.99 y .97) para la categoría 0 (errores) y 1.01, 1.05 para la categoría 1 (aciertos).

Las columnas M ->C y C ->M, al extremo de la tabla presentan la coherencia de la prueba con la muestra, a partir de los porcentajes esperados y los obtenidos en las medidas, para cada categoría de respuesta. Los datos indican que los valores observados son cercanos a los esperados.

La columna M->C representa la coherencia del examen con la muestra, expresada por el porcentaje real de las medidas que se esperaba que produjeran las observaciones en esta categoría. En el presente caso se observa una coherencia razonable entre la expectativa y los datos observados.

La columna C->M representa el porcentaje de las observaciones que se produjeron por medidas que corresponden a la categoría. De manera similar, en el presente caso se observa una coherencia razonable entre la expectativa y los datos observados.

Enseguida se presenta la Figura 6.2 con la curva dicotómica de probabilidad, que es una predicción de la forma en la que se comportaría el examen en muestras de sustentantes similares.

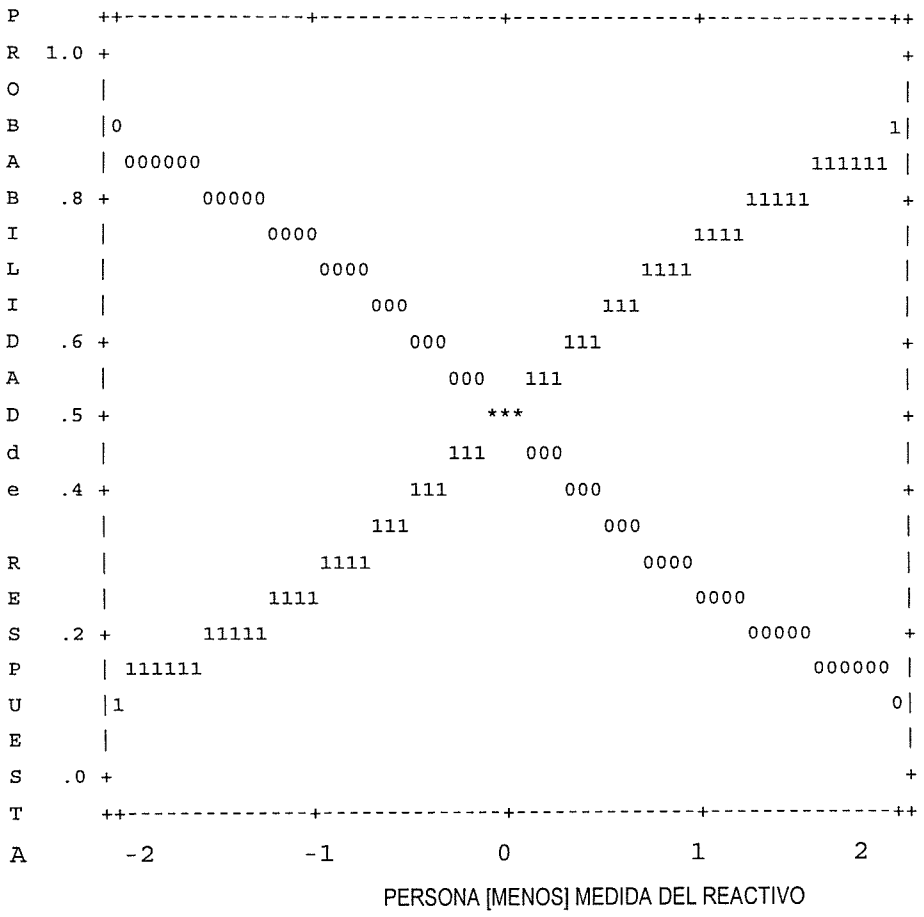



Figura 6.2. Curvas dicotómicas de probabilidades de respuesta.

En la figura puede notarse que la categoría 0 (respuesta incorrecta) tendrá una probabilidad de -2 lógitos a .90 de respuesta calibrada. De igual manera, la categoría 1 (respuesta correcta) tendrá una probabilidad de .90 a +2 lógitos de respuesta calibrada.

De manera inversa, a -2 lógitos de habilidad la respuesta correcta tiene .10 de probabilidad, mientras que a +2 lógitos de habilidad, la respuesta correcta tiene .90 de probabilidad de ocurrir. En la habilidad media (0 lógitos), las categorías 1 y 0 tienen .50 de probabilidad de ocurrir.

A continuación se presenta la Figura 6.3 que muestra la distribución de la Medida con respecto a la dificultad en la muestra analizada.

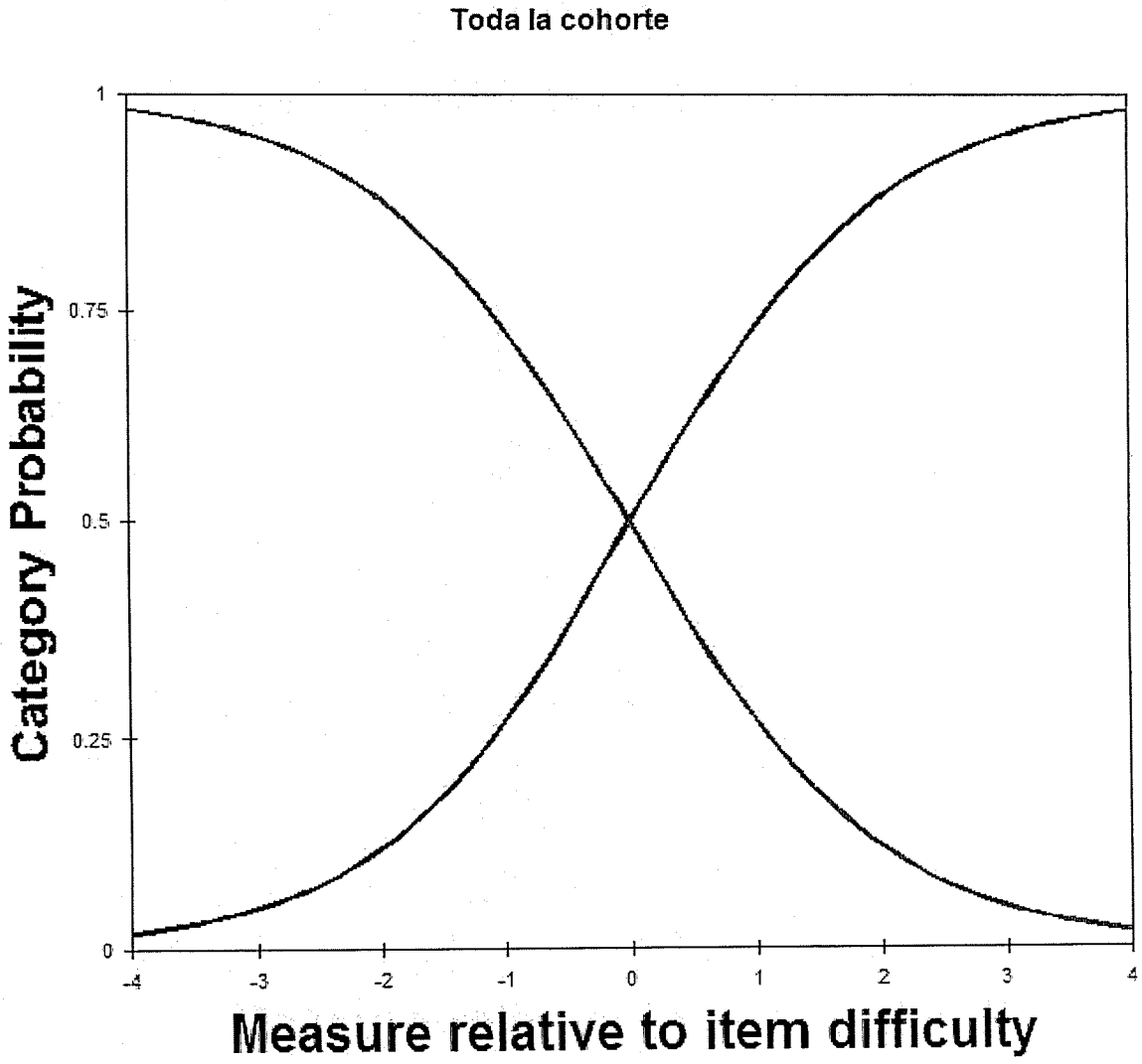



Figura 6.3 Probabilidades de respuesta en relación con la dificultad de los reactivos para la muestra analizada (figura obtenida del archivo de salida).

Se observan las siguientes características:

La línea roja representa la probabilidad de respuesta incorrecta. Como puede verse, la línea es decreciente a medida que aumenta la habilidad de los sustentantes.

La línea azul representa la probabilidad de respuesta correcta en forma creciente, a medida que aumenta la habilidad en lógitos.

Estos resultados pueden considerarse indicadores del comportamiento del examen. La Figura 6.4 muestra la curva característica del examen.

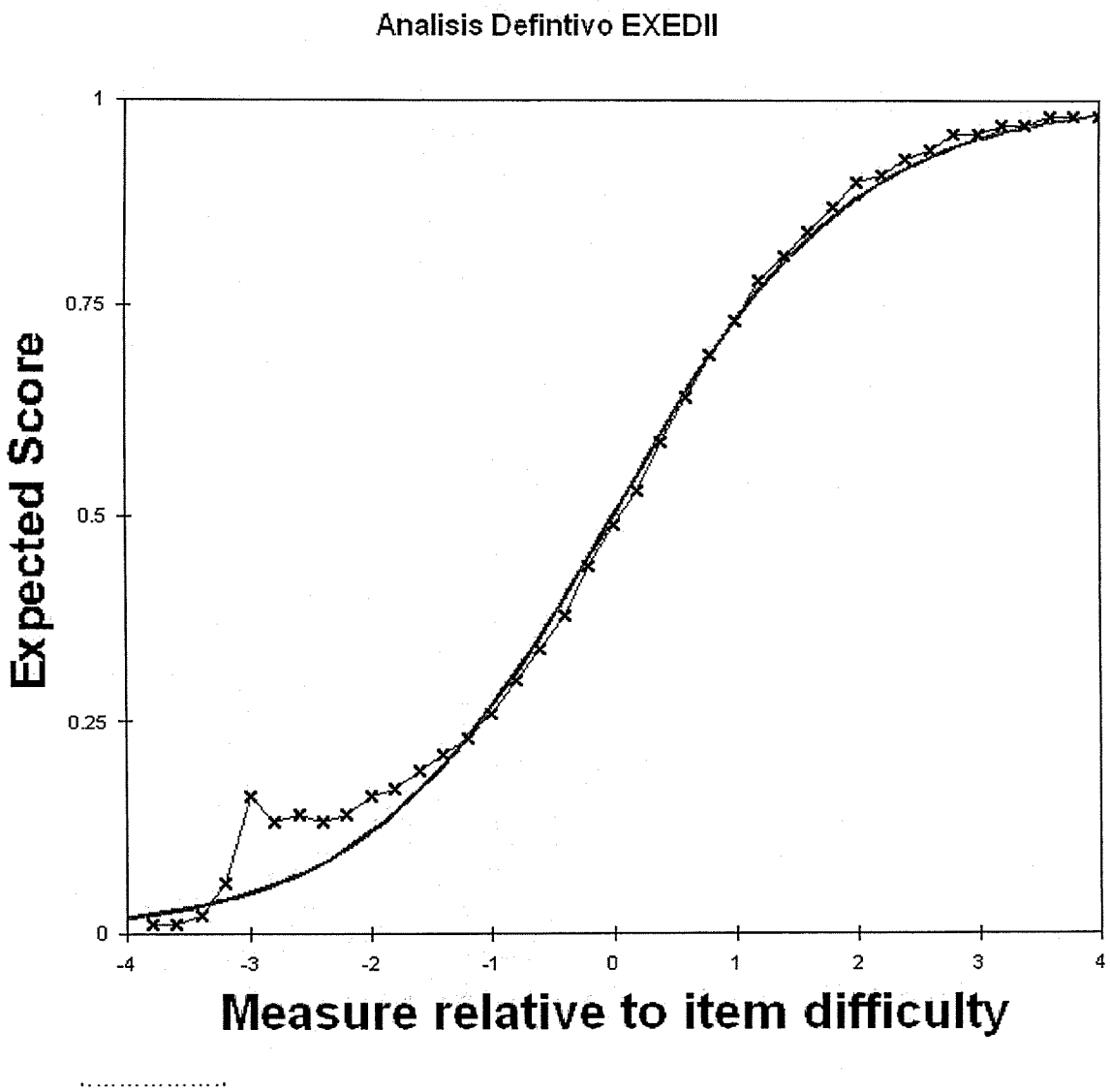



Figura 6.4. Ajuste de los datos observados con el modelo (figura obtenida del archivo de salida).

Esta gráfica representa la congruencia entre los patrones de respuesta observados en la muestra analizada y las probabilidades de respuesta esperadas según el modelo. Se observa que los patrones de respuesta coinciden con las expectativas del modelo, expresadas como probabilidades, entre -4.0 y 4.0 lógitos de habilidad calibrada, con excepción de un desajuste alrededor de -3 lógitos.

## 6.3. Estructura factorial

Todos los estudios que se han hecho para conocer el comportamiento del EXEDII se han realizado desde el marco de referencia de la TCM. Al diseñar la estrategia de validación del constructo del EXEDII se consideró conveniente realizar análisis factorial (AF) para identificar las dimensiones latentes del examen, toda vez que el AF es un método para representar las relaciones entre un grupo de variables, en términos de un número reducido de variables hipotéticas (Spearritt, 1988; Thompson, 2005).

La literatura sobre el tema abunda en ejemplos de AF, o de análisis de componentes principales realizados indistintamente en variables binarias y graduadas, pero algunos autores alertan sobre la poca credibilidad de esos resultados (Nunally y Bernstein, 1995; Ferrando, 1996). El problema es que para obtener resultados que tengan credibilidad en un AF es necesario que los datos utilizados provengan de variables continuas y multivariadas, distribuidas normalmente y, ni todos los estudios pueden demostrar esos supuestos, ni todos los procedimientos son suficientemente rigurosos como para demostrar hasta qué punto los reactivos que componen una escala miden la misma dimensión y no otra (Ferrando, 1996).

### 6.3.1. Indagación de los patrones de convergencia.

En esta estrategia de búsqueda de evidencias de constructo del EXEDII se requiere investigar si las variables observadas --las respuestas a los reactivos— pueden agruparse y comprenderse en un número menor de variables no observadas o factores. El Análisis Factorial (AF) trabaja sobre la correlación de las variables observadas, en el que los coeficientes de correlación expresan la relación concomitante entre dos o más variables. Existen diferentes métodos de correlación, dependiendo del tipo de datos observados. Por ejemplo, el más frecuente en la literatura es el coeficiente de Pearson, que es indicado cuando se correlacionan dos variables continuas cuya relación se asume lineal. Otro método frecuentemente reportado es la correlación biserial y la biserial puntual, que son indicadas cuando una variable es continua y la otra dicotómica. El método de Spearman se debe utilizar para variables ordinales. En el caso

de tener variables continuas, dicotomizadas con algún punto de corte y distribuidas normalmente, como es el caso del EXEDII, debe utilizarse la correlación tetracórica (Tristán y Vidal, 2006).

Con base en lo anterior, el AF de variables dicotómicas no debe hacerse sobre la matriz de correlaciones de Pearson, sino sobre la matriz inter-item de correlaciones tetracóricas, porque la configuración de cada una de ellas parte de suposiciones distintas. En la correlación tetracórica, al dicotomizar las variables se obtienen cuatro celdas en una tabla de contingencia, que corresponden a todas las combinaciones posibles de grupos superiores e inferiores de las dos variables (Guilford 1984; citado en Tristán y Vidal, 2006).

### 6.3.2. Análisis Factorial Exploratorio.

La lógica para el AFE se basa en una estrategia frecuentemente utilizada en el análisis de pruebas psicológicas y educativas, en las que se espera una estructura bifactorial que evidencia un factor general y uno o más factores. Esta estructura resulta de la restricción que establece que cada reactivo tenga una carga mayor a cero en la dimensión primaria y además en uno de los otros grupos o factores (Gibbons, Bock y Hedeker, 2007). Para fines del AFE se siguió la estrategia utilizada en la tesis doctoral de González Montesinos (2004) Se conformó un archivo de datos con todos los resultados de los sustentantes del EXEDII de las cohortes de 2005 y 2006. Los patrones de respuesta de cada sustentante son dicotómicos de manera que se tiene que: acierto=1 y error=0.

El AFE se realizó con ayuda del programa de cómputo TESTFACT 4 (Bock, Gibbons, Schilling, Muraki, Wilson y Wood, 2003). Este programa utiliza el método denominado Análisis Factorial de Información Completa (*Full Information Factor Analysis, FIFA* por sus siglas en inglés) porque hace uso de los procedimientos de la Teoría de respuesta al Reactivo tomando en cuenta toda la información disponible en el patrón de respuesta de cada sustentante. El procedimiento que se siguió es el de Máxima Verosimilitud Marginal que lleva al máximo posible las cargas factoriales a partir de los patrones de respuesta observados (Bock, Gibbons y Muraki,

1998). El análisis no sigue una estrategia estándar sino que se requirió de tomar una serie de decisiones en cada paso a calcular, dependiendo de los objetivos del análisis. Para efectuar el AFE se elaboró el archivo de instrucciones que se presenta en el Anexo 5.2, el cual está basado en el archivo de instrucciones de González Montesinos (2004).

### 6.3.3. Resultados

El archivo de salida del AFE se muestra en el Anexo 6.2. En este capítulo se destacan los resultados más importantes para el propósito del estudio. Una vez que hubo "corrido" el análisis se verificó que hubiera llegado a una solución y se procedió a hacer el análisis del archivo de salida (se muestra completo en el Anexo 6.2) destacando los aspectos que se describen a continuación:

Dado que inicialmente el programa proporciona los estadísticos básicos, en el archivo de salida se observa que el puntaje promedio es de 57, con desviación estándar de 20.26. La consistencia interna o confiabilidad arroja un valor de 0.956. (Kuder-Richardson KR20).

La revisión del histograma muestra que graficando en el eje de las X los puntajes y en el eje de las Y las frecuencias de éstos se dibuja una curva semejante a la normal, con varios picos correspondientes a un amplio rango de puntajes desde 20 hasta 98.

En la matriz de correlación se observan cuatro raíces latentes positivas con valores de: 32.098; 3.891; 2.891 y 2.569.

Después de la rotación Varimax se obtuvieron tres factores. Enseguida se presenta la Tabla 6.4 que muestra los factores obtenidos y sus cargas factoriales después de la rotación. Los factores fueron nombrados de acuerdo con su carga factorial y el contenido conceptual. En la Tabla 6.4 los reactivos aparecen escritos en itálicas cuando presentan cargas interpretables en el Factor 1, con negritas cuando cargan en el Factor 2 y con un tipo de letra diferente cuando presentan carga interpretable en el Factor 3.

Tabla 6.4. Factores obtenidos después de la rotación Varimax.

| Reactivos | Conocimiento de de la lengua | Conocimiento organizacional | Conocimiento textual |
|---|---|---|---|
| 1 ITEM1 | *0.404* | 0.238 | 0.185 |
| 2 ITEM2 | *0.602* | 0.188 | 0.215 |
| 3 ITEM3 | *0.509* | 0.258 | 0.190 |
| 4 ITEM4 | *0.343* | 0.146 | 0.199 |
| 5 ITEM5 | *0.620* | 0.243 | 0.221 |
| 6 ITEM6 | *0.611* | 0.188 | 0.234 |
| 7 ITEM7 | *0.568* | 0.296 | 0.309 |
| 8 ITEM8 | *0.541* | 0.109 | 0.099 |
| 9 ITEM9 | 0.140 | 0.043 | -0.003 |
| 10 ITEM10 | *0.636* | 0.213 | 0.309 |
| 11 ITEM11 | *0.638* | 0.265 | 0.278 |
| 12 ITEM12 | 0.284 | 0.060 | 0.138 |
| 13 ITEM13 | *0.572* | 0.108 | 0.270 |
| 14 ITEM14 | 0.273 | 0.220 | 0.124 |
| 15 ITEM15 | *0.576* | 0.209 | 0.280 |
| 16 ITEM16 | 0.060 | 0.060 | 0.092 |
| 17 ITEM17 | *0.497* | 0.195 | 0.250 |
| 18 ITEM18 | *0.648* | 0.215 | 0.238 |
| 19 ITEM19 | *0.608* | 0.213 | 0.274 |
| 20 ITEM20 | *0.383* | 0.182 | 0.235 |
| 21 ITEM21 | *0.411* | 0.133 | 0.136 |
| 22 ITEM22 | *0.518* | 0.184 | 0.250 |
| 23 ITEM23 | *0.564* | 0.167 | 0.243 |
| 24 ITEM24 | *0.623* | 0.284 | 0.275 |
| 25 ITEM25 | *0.628* | 0.133 | 0.226 |
| 26 ITEM26 | *0.576* | 0.225 | 0.348 |
| 27 ITEM27 | *0.697* | 0.244 | 0.303 |
| 28 ITEM28 | *0.494* | 0.146 | 0.170 |
| 29 ITEM29 | *0.563* | 0.206 | 0.238 |
| 30 ITEM30 | *0.724* | 0.259 | 0.281 |
| 31 ITEM31 | *0.660* | 0.165 | 0.232 |
| 32 ITEM32 | *0.696* | 0.189 | 0.255 |
| 33 ITEM33 | 0.068 | **0.653** | 0.169 |
| 34 ITEM34 | 0.131 | 0.298 | 0.090 |
| 35 ITEM35 | 0.231 | **0.338** | 0.217 |
| 36 ITEM36 | *0.352* | **0.385** | 0.321 |
| 37 ITEM37 | *0.349* | **0.512** | 0.232 |
| 38 ITEM38 | 0.164 | **0.430** | 0.188 |
| 39 ITEM39 | *0.392* | **0.473** | 0.215 |
| 40 ITEM40 | *0.460* | **0.431** | 0.228 |
| 41 ITEM41 | 0.202 | **0.340** | 0.065 |
| 42 ITEM42 | *0.312* | **0.309** | 0.166 |
| 43 ITEM43 | *0.390* | **0.386** | 0.285 |
| 44 ITEM44 | *0.376* | **0.376** | 0.304 |
| 45 ITEM45 | *0.542* | **0.533** | 0.359 |
| 46 ITEM46 | *0.512* | **0.464** | 0.368 |
| 47 ITEM47 | 0.229 | **0.801** | 0.205 |
| 48 ITEM48 | 0.263 | **0.736** | 0.217 |
| 49 ITEM49 | 0.265 | **0.774** | 0.230 |
| 50 ITEM50 | 0.140 | **0.829** | 0.208 |
| 51 ITEM51 | 0.019 | **0.668** | 0.107 |

| | | | | |
|---|---|---|---|---|
| 52 | ITEM52 | -0.016 | **0.380** | 0.069 |
| 53 | ITEM53 | 0.101 | **0.531** | 0.182 |
| 54 | ITEM54 | *0.329* | **0.441** | 0.234 |
| 55 | ITEM55 | 0.148 | **0.455** | 0.147 |
| 56 | ITEM56 | 0.189 | **0.481** | 0.180 |
| 57 | ITEM57 | *0.371* | **0.586** | 0.281 |
| 58 | ITEM58 | 0.260 | **0.458** | 0.184 |
| 59 | ITEM59 | *0.500* | **0.538** | 0.332 |
| 60 | ITEM60 | *0.396* | **0.500** | 0.278 |
| 61 | ITEM61 | 0.243 | **0.345** | 0.252 |
| 62 | ITEM62 | 0.160 | 0.203 | 0.079 |
| 63 | ITEM63 | 0.238 | **0.373** | 0.178 |
| 64 | ITEM64 | 0.265 | **0.416** | 0.229 |
| 65 | ITEM65 | *0.434* | **0.506** | 0.264 |
| 66 | ITEM66 | *0.338* | **0.358** | 0.219 |
| 67 | ITEM67 | 0.291 | 0.242 | 0.282 |
| 68 | ITEM68 | *0.506* | 0.264 | 0.388 |
| 69 | ITEM69 | 0.234 | 0.211 | 0.132 |
| 70 | ITEM70 | 0.201 | 0.182 | 0.243 |
| 71 | ITEM71 | 0.292 | 0.167 | 0.195 |
| 72 | ITEM72 | *0.334* | 0.277 | 0.254 |
| 73 | ITEM73 | *0.403* | 0.201 | 0.305 |
| 74 | ITEM74 | 0.291 | **0.354** | 0.260 |
| 75 | ITEM75 | *0.423* | 0.287 | 0.302 |
| 76 | ITEM76 | 0.063 | 0.153 | 0.190 |
| 77 | ITEM77 | 0.295 | 0.291 | 0.225 |
| 78 | ITEM78 | 0.119 | 0.270 | 0.275 |
| 79 | ITEM79 | 0.235 | 0.122 | 0.235 |
| 80 | ITEM80 | *0.331* | 0.183 | **0.346** |
| 81 | ITEM81 | 0.198 | 0.182 | 0.233 |
| 82 | ITEM82 | *0.369* | **0.329** | 0.545 |
| 83 | ITEM83 | *0.409* | 0.296 | 0.383 |
| 84 | ITEM84 | *0.390* | 0.183 | 0.480 |
| 85 | ITEM85 | 0.247 | 0.270 | 0.725 |
| 86 | ITEM86 | 0.244 | 0.066 | 0.235 |
| 87 | ITEM87 | 0.413 | 0.240 | 0.479 |
| 88 | ITEM88 | 0.267 | 0.185 | 0.411 |
| 89 | ITEM89 | 0.166 | 0.182 | 0.456 |
| 90 | ITEM90 | 0.187 | 0.176 | 0.593 |
| 91 | ITEM91 | 0.249 | 0.163 | 0.450 |
| 92 | ITEM92 | *0.385* | **0.326** | 0.735 |
| 93 | ITEM93 | *0.335* | 0.262 | 0.641 |
| 94 | ITEM94 | 0.265 | 0.210 | 0.562 |
| 95 | ITEM95 | *0.335* | 0.213 | 0.664 |
| 96 | ITEM96 | 0.121 | 0.259 | 0.562 |
| 97 | ITEM97 | 0.256 | 0.179 | 0.502 |
| 98 | ITEM98 | 0.254 | 0.264 | 0.752 |
| 99 | ITEM99 | *0.342* | 0.204 | 0.611 |
| 100 | ITEM100 | *0.307* | 0.209 | 0.580 |

De acuerdo con la literatura las cargas mínimas en los factores deben rebasar el .30 para considerarse interpretables (por ejemplo, Bock, Gibbons y Muraki,1988; Nunnally y

Bernstein,1995). Con base en ese criterio se puede observar que los tres factores aglutinan cada uno a un grupo de reactivos diferente: el primer factor contiene a la mayor parte de los reactivos, particularmente los que corresponden a la subescala de Comprensión auditiva (1-32). Sin embargo, reactivos de las otras dos subescalas cargan en ese factor. Por esa razón se interpreta que F1 es una variable que representa una habilidad de manejo general del idioma, que en términos de Bachman y Palmer (1996) podría estar representando la habilidad de Conocimiento de la lengua. El segundo factor contiene a la mayoría de los reactivos de la subescala de Gramática (33-66). Por ello se ha denominado a este factor Conocimiento textual, que de acuerdo con esos autores es uno de los componentes de la habilidad general. Finalmente, los reactivos que se agrupan en el tercer factor corresponden en su mayoría a la subescala de Lectura, por lo que se denominó a este factor Conocimiento textual porque de acuerdo con la teoría, esta variable estaría representando al componente de la habilidad general que está relacionado con la comprensión y producción de textos, sean hablados o escritos.

Después de la rotación, la varianza explicada por cada uno de los factores también es congruente con la teoría sustantiva: el primer factor explica el 32.16 de la varianza; el segundo explica el 3.52 y el tercero explica el 2.28. Juntos, los tres factores explican el 37.95% de la varianza

Es de notar que no todos los reactivos presentaron cargas interpretables en algún factor, ya que el 15% no presentó cargas interpretables en ninguno de los tres factores extraídos. Así, el Factor 1 alberga al 34% de los reactivos, el Factor 2 incluye al 28% de los reactivos y el Factor 3 contiene al 19% de los reactivos.

El siguiente paso del AFE consiste en realizar una prueba para investigar si el número de factores postulado en el primer análisis es el que más adecuadamente explica la estructura factorial del instrumento evaluado. Para ello fue necesario volver a "correr" el análisis con un número distinto de factores y verificar la mejor solución. Se procedió de la siguiente manera:

Considerando que el diseño del examen postula tres factores, la hipótesis inicial establece que:

H0: Un modelo de 3 factores permite una descripción adecuada de los datos.

Se plantea una hipótesis alternativa que establece:

H1: Un modelo de 2 factores permite una descripción adecuada de los datos.

Es decir, se volvió a realizar un AFE hipotetizando dos factores. Los resultados obtenidos presentan los siguientes valores:

CHI cuadrada = 204026.33 DF = 1960.00 P = 0.000

Con el propósito de probar H0 contra H1 se calculó la diferencia de las X2 y de los grados de libertad en ambos análisis.

Dado que:
X2 de H0 = 204026.33 y X2 de H1 = 202256.00
la diferencia = 1770
con 1960 – 1862 = 98 grados de libertad.

Dado que en las tablas el valor crítico de X2 (100) = 124.342 al nivel de $p<.05$, se rechaza H1 y se acepta H0 concluyendo que un modelo de tres factores permite una descripción adecuada de la estructura del EXEDII.

Con base en el Análisis Factorial Exploratorio aplicado a las variables conformadas por los reactivos del EXEDII se concluye que existe evidencia de que el instrumento analizado en efecto mide el constructo definido como "Manejo del inglés como lengua extranjera al nivel intermedio, en estudiantes que egresan de la UABC" ya que las cien variables o reactivos que constituyen el instrumento se pudieron reducir a tres factores subyacentes, lo cual es congruente con el contenido conceptual de los reactivos y con el diseño del instrumento. Se obtiene esta conclusión con base en los siguientes criterios: 1) las variables presentan carga significativa (.30

ó mayor) en cada factor retenido; 2) las variables que cargan en cada factor comparten un significado conceptual; 3) las variables que presentan cargas interpretables en factores diferentes describen constructos conceptualmente diferentes; 4) el patrón de carga presenta "estructura simple". Adicionalmente, en la prueba de bondad de ajuste se obtiene un nivel de significancia mayor a 0.05, por lo que se número menor de reactivos con cargas factoriales interpretables acepta la hipótesis nula que establecía que "los tres factores extraídos son suficientes para explicar la estructura factorial del instrumento valorado. El Factor 1 agrupa al 87.5% de los reactivos de la subescala de Comprensión auditiva y es el factor que explica mayor proporción de varianza (32.16%). El Factor 2 agrupa al 82.3% de los reactivos de Gramática y explica el 3.52% de la varianza. El Factor 3 agrupa al 53% de los reactivos de la subescala de lectura, explicando el 2.28% de la varianza. El 15% de los reactivos no presentaron cargas interpretables en ningún factor. De ellos, el 12.5% corresponde a Comprensión auditiva, el 2.94% a Gramática y el 29.4% de Lectura.

Hasta aquí se presentan los resultados obtenidos para la prueba en su conjunto. En el siguiente apartado se revisan los datos obtenidos por reactivo, tanto en el Análisis de Rasch, como en el AFE con el propósito de integrar la información obtenida y posteriormente interpretar los hallazgos de la estrategia para recabar evidencias de validez de constructo.

## 6.4. Integración de los resultados del Análisis de Rasch y Análisis Factorial Exploratorio.

Desde el punto de vista de la prueba en su conjunto los resultados de los dos análisis arrojan información referente a la dimensionalidad y la estructura factorial de la misma. Pero también para cada uno de los reactivos se obtuvieron datos que muestran su ajuste al modelo Rasch y sus cargas factoriales.

Teóricamente se esperaría que todos los reactivos del EXEDII midieran una misma dimensión, lo que en términos del Análisis de Rasch equivaldría a que todos los reactivos se ajustaran al modelo. Por otra parte desde el punto de vista de la estructura del examen se

esperaría que en el Análisis factorial todos los reactivos de cada subescala presentaran cargas altas en un factor, así como cargas cercanas a cero en los demás factores.

Para interpretar los resultados obtenidos en los dos análisis realizados fue necesario apegarse a los criterios que se mencionan en la literatura. Por ejemplo, en los estadígrafos de ajuste interno y externo, aquellos valores observados que rebasen los rango de 1 a +1.3 en la media cuadrática y -2 a +2 en los estadígrafos estandarizados se consideran que no se ajustan al modelo (Linacre, 1991-2006; González Montesinos, 2008).

Con respecto a la carga factorial de los reactivos individuales se tomaron en cuenta las recomendaciones de los autores del software utilizado para AFE (Bock, Gibbons y Muraki, 1988) y de otros autores teóricos, o que han realizado estudios similares (Kachigan, 1991; Nunnally, 1995; Ferrando, 1996) según los cuales la carga mínima interpretable en un factor es de .30 ó .35, dependiendo de otros elementos de juicio que el investigador juzgue pertinentes. La Tabla 6.7 integra los reactivos que no se ajustaron al modelo Rasch, que no presentaron cargas factoriales interpretables, o ambas condiciones.

Tabla 6.5 Reactivos que presentan anomalías en los indicadores de ajuste al modelo Rasch y/o que no presentan cargas factoriales interpretables en ningún factor.

| | Factores | | | Ajuste interno (Infit) | | Ajuste externo (Outfit) | |
|---|---|---|---|---|---|---|---|
| Reactivos | F1 | F2 | F3 | MNSQ 1 a 1.3 | ZSTD -1 a +2 | MNSQ 1 a 1.3 | ZSTD -1 a +2 |
| 16 | .60 | .060 | .092 | | 2.1 | 1.75 | 2.2 |
| 62 | .160 | .203 | .079 | 1.31 | 2.1 | 1.45 | 2.4 |
| 76 | .063 | .153 | .190 | 1.34 | 2.6 | 1.41 | 2.3 |
| 52 | * | .380 | * | 1.34 | 2.5 | 1.45 | 2.5 |
| 59 | .500 | .538 | .332 | | -2.2 | | -2.2 |
| 30 | .724 | * | * | | -2.1 | | |
| 12 | .284 | .060 | .138 | | | 1.40 | |
| 34 | .131 | .298 | .090 | | | 1.33 | |
| 9 | .140 | .430 | -.003 | | | 2.42 | |
| 41 | .340 | * | * | | | 1.38 | |
| 70 | .201 | .182 | .243 | | | 1.32 | |
| 79 | .235 | .122 | .235 | | | 1.39 | |
| 86 | .244 | .066 | .235 | | | 1.43 | |
| 69 | .234 | .211 | .132 | | | | |
| 14 | .273 | .220 | .124 | | | | |
| 67 | .291 | .242 | .282 | | | | |
| 71 | .292 | .167 | .195 | | | | |
| 77 | .295 | .291 | .225 | | | | |
| 78 | .119 | .122 | .235 | | | | |
| 81 | .198 | .182 | .233 | | | | |

*No hay carga interpretable.

En la tabla anterior, es posible observar que los reactivos 16, 62 y 76 no presentan cargas interpretables y además no se ajustan al modelo Rasch. Enseguida, el reactivo 52 rebasa los valores esperados para los cuatro indicadores de ajuste interno y externo pero presenta carga factorial en F2 (congruente con su contenido conceptual). El reactivo 59 presenta carga factorial interpretable en los tres factores, pero se muestra elevado el estadígrafo estandarizado de ajuste interno. Siguiendo con la lista, el reactivo 30 carga en F1 (congruente con su contenido conceptual) pero rebasa el criterio de ajuste interno estandarizado. Los reactivos 12 y 34 no presentan carga interpretable y además rebasan el criterio de ajuste externo en el estadígrafo no estandarizado. El reactivo 9 no presenta carga factorial interpretable y el valor de la media cuadrática de ajuste externo es elevado. El reactivo 41 tiene carga factorial interpretable en F1, pero no es congruente con su contenido conceptual y el estadígrafo de ajuste externo se observa fuera del rango esperado. Los reactivos 70, 79 y 86 tienen en común que no presentan carga factorial interpretable y presentan valores fuera del criterio en el estadígrafo no

estandarizado de ajuste externo. Los reactivos 69, 14, 67, 71, 77, 78 y 81 presentan una situación similar, ya que no cargan en ningún factor en el AFE, pero se ajustan al modelo Rasch.

Llama la atención que el 30% de los reactivos del examen cargan en dos factores y en el caso de siete de ellos, en tres factores. En la Discusión se retoma este punto que podría ser un indicador de anomalía, puesto que se esperaría que las cargas factoriales presentaran estructura simple para ser interpretables, o bien señalaría una habilidad general y dos componentes de la habilidad.

**Recapitulando:**

Los resultados obtenidos en la indagación de validez de constructo indican que el EXEDI:

- Es una escala que mide una dimensión y sus reactivos constituyen un conjunto de elementos que funcionan en una misma dirección.
- Presenta consistencia interna ya que el valor estimado de confiabilidad es de .95 (Modelo Rasch) y .956 (AFE) en AFE con la fórmula de Kuder-Richardson.
- Presenta estabilidad en la medición pues los errores de los reactivos son bajos (alrededor de .05).
- Consiste de un grupo de reactivos que, en su mayoría se ajustan al modelo Rasch.
- Contiene reactivos que representan todo el rango de dificultad intermedia.
- Mide una habilidad general de manejo del idioma y dos habilidades secundarias que valoran el conocimiento y uso de reglas gramaticales y habilidades de comprensión de textos escritos y lenguaje hablado, lo cual es congruente con la teoría sustantiva.
- Algunos reactivos requieren ser revisados minuciosamente para decidir si se quedan en la prueba tal y como están, se modifican o se eliminan.

# 7. DISCUSION Y CONCLUSIONES

En este capítulo se discuten los resultados obtenidos en las dos vertientes de que consta la investigación de las evidencias de validez del EXEDII. Se discuten los hallazgos a la luz del concepto de la Utilidad de las pruebas, que es el indicador más completo de la validez de un examen de acuerdo con los fundamentos teóricos de esta tesis.

En el capítulo 4 se plantea la aproximación de Bachman y Palmer (1996) a la enseñanza y evaluación de los idiomas, así como a la construcción y validación de las pruebas de lenguas. De acuerdo con estos autores los instrumentos que evalúan el manejo de idiomas permiten colectar información útil que beneficia a una variedad de individuos, pero se requiere demostrar que los puntajes obtenidos en los exámenes son confiables y que la manera en la que se interpretan esos puntajes es válida.

De acuerdo con Bachman (2004) los resultados de las pruebas de lenguas pueden ser entendidos desde dos contextos: el de la lingüística aplicada y el de la medición. En el primero se incluye la temática de la naturaleza del uso del lenguaje y de sus tareas, el aprendizaje y la habilidad adquirida. El segundo tiene que ver con la relación entre los resultados cuantitativos por un lado y su significado, interpretación y uso por el otro. Por ello la demostración de la utilidad de una prueba deberá incluir la investigación empírica de la ejecución de los sustentantes en la prueba.

Derivado de lo anterior las indagaciones pueden enfocarse hacia: a) los procesos o las estrategias que utilizan los sustentantes para responder y b) los puntajes obtenidos por los sustentantes. La literatura sobre la evaluación de los idiomas abunda en investigaciones que atienden a uno o al otro enfoque, o incluso ambos, pero el sentido de las interpretaciones sobre los resultados sigue siendo el responder a las preguntas de investigación. En este estudio se recurre al concepto de utilidad de la prueba, el cual constituye un eje que permite articular y dar sentido a los hallazgos de la investigación que se reporta en esta tesis.

En el apartado 4.5.3.4. se planteó el concepto de la Utilidad de las pruebas que según Bachman y Palmer (1996) puede evaluarse mediante seis indicadores: confiabilidad, validez de constructo, autenticidad, interactividad, impacto y viabilidad. En los siguientes apartados se exponen los hallazgos de la presente investigación desde el punto de vista de su aporte a la utilidad del EXEDII.

## 7.1 Confiabilidad.

La confiabilidad de una prueba se refiere a la consistencia de la medición y puede evaluarse de distintas formas. Derivadas de la estrategia de validación de este estudio las evidencias de confiabilidad provienen de dos fuentes: los juicios que los expertos hacen acerca de tres aspectos que en la literatura son mencionados reiteradamente como indicadores de confiabilidad y de validez (relevancia, representatividad y pertinencia de los reactivos) y los análisis estadísticos que estiman las correlaciones de los patrones de respuesta de los sustentantes. Enseguida se presentan los hallazgos de la investigación que aportan evidencias de confiabilidad del EXEDII:

1. Los resultados cuantitativos de la apreciación de los jueces indica que la mayoría de las tareas que expresan los nodos del EXEDII son relevantes, representativas y pertinentes, lo que aporta evidencia de la confiabilidad del examen al mostrar la consistencia interna del instrumento. En esta investigación se consideran como evidencias de validez de contenido los consensos y los acuerdos de al menos el 75% de las opiniones de los jueces, en el sentido de que el nodo y sus características coinciden con las del criterio (DULM). Por ello, a partir de las respuestas de los jueces se considera que la mayor parte de los reactivos del EXEDII están representados en la conceptualización de lo que debería medir el EXEDII. No obstante, los casos en los que no hubo consenso, ni coincidencia igual o mayor a la del criterio se presentan en la Tabla 7.1. Así, se observa que el 97.3% de los nodos del examen fueron calificados por los jueces como relevantes y representativos, con excepción de dos reactivos de Comprensión auditiva (6.3% de esa subescala). De manera similar la pertinencia del total de los nodos es del 97% salvo tres reactivos de la misma escala (9.4% de los reactivos de Comprensión auditiva) que fueron evaluados negativamente. Finalmente la apreciación de la claridad en la descripción de los nodos se investigó para recabar la opinión de los jueces a este respecto, en caso de que la forma en la que se enuncia el nodo presentara dificultades para comprender la tarea y en este sentido se obtiene que el 95% se consideraron claramente definidos.

El hecho de que los jueces consensual o mayoritariamente califiquen a la mayor parte de los reactivos en un sentido determinado habla de la congruencia entre lo que mide cada reactivo en relación con lo que miden los demás.

2. La evaluación de las características de los reactivos también proporciona información que aporta evidencias de confiabilidad. La Tabla 7.1 también muestra los porcentajes de los casos que no obtuvieron consenso o al menos el 75% de acuerdo entre los jueces al respecto de las características. Así, el 93% de

los reactivos obtuvieron consenso de medir el nivel apropiado, con excepción de tres reactivos de comprensión auditiva (9.4%) y cuatro de Lectura (11.8% de la subescala). Algo similar puede afirmarse de la apreciación de los expertos con respecto a que las características de los reactivos reflejan las características de los sustentantes, ya que el 96% de ellos obtuvieron consenso en este sentido, exceptuando a tres reactivos de comprensión auditiva (9.4%) y uno de Lectura 2.9% de la subescala). Sin embargo, la apreciación de que los reactivos reflejan el contexto cultural de los sustentantes es distinta de las demás características ya que las tres subescalas obtienen menores proporciones de consenso o acuerdo de los jueces. La subescala de Comprensión auditiva cuenta con siete reactivos calificados negativamente, mientras que la de Gramática obtiene diez y la de Lectura nueve, sumando entre las tres el 26% del total de reactivos del examen. Finalmente la apreciación de que las características de los reactivos pudieran estar produciendo sesgo obtiene consenso en el 26% de los reactivos de Lectura, que en el total de la prueba significa el 9%. Esta última característica se interpreta de manera diferente a las anteriores porque el que un reactivo produzca sesgo pone en duda su validez de contenido. Las razones por las que los jueces manifestaron la posibilidad de sesgo se refieren a que algunos reactivos tocan temas que pudieran incomodar a algunos sustentantes, tales como obesidad o preferencias musicales.

Entonces, el acuerdo en las opiniones de los jueces con respecto a la evaluación de las características de los reactivos y las competencias evaluadas aporta evidencias de confiabilidad de la prueba, aunque se detecta un aspecto de los reactivos que requiere revisión y posiblemente modificación de algunos reactivos. Este aspecto se retoma en las recomendaciones que se hacen para mejorar el EXEDII.

3. Las evidencias de confiabilidad que aportan los análisis estadísticos se exponen en seguida. El Análisis de Rasch proporciona evidencias de: **a)** consistencia interna del instrumento al resultar positiva la dimensionalidad lo que aporta una prueba de que los reactivos constituyen un conjunto de elementos que funcionan en una misma dirección; **b)** el valor estimado de confiabilidad a través de la fórmula de Kuder-Richardson, que es de .95 considerado un valor alto en la literatura; **c)** el error de medición promedio resulta en .05 respecto de los reactivos que es un valor bajo, lo que significa una mayor confianza en que se está midiendo un atributo de manera estable; **d)** el valor del error del modelo (model RMSEA) se interpreta como el límite superior de los estimados de confiabilidad y en el presente caso los valores del modelo y los valores observados coinciden, tanto para los reactivos (.05), como para los sustentantes (.26); **e)** la mayoría de los reactivos se ajustan al modelo, lo que indica que sus características son consistentes.

4. El Análisis Factorial Exploratorio realizado aporta las siguientes evidencias de confiabilidad, derivadas de la naturaleza del análisis, es decir la correlación entre variables: **a)** el valor estimado de confiabilidad es de .956 utilizando la fórmula de Kuder-Richardson, que es la más indicada para reactivos dicotómicos; **b)** el hallazgo de factores constituye una evidencia de confiabilidad porque la asociación de cada reactivo con el factor implica una relación entre variables.

Tabla 7.1. Acuerdos menores a 75% entre los panelistas: aspectos evaluados de los Nodos y Características de los reactivos, expresados en frecuencias relativas y porcentajes. Subescala: Lectura

| | NODOS | | | | CARACTERISTICAS DE LOS REACTIVOS | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Relevante | Representativo | Pertinente | Claramente descrito | Similar a realidad | Nivel apropiado | Refleja características sustentantes | Refleja contexto cultural sustentantes | Provoca sesgo |
| | FR<br>% | FR<br>% | FR<br>% | FR<br>% | FR<br>% | FR<br>% | FR<br>% | FR<br>% | FR<br>% |
| Comprensión auditiva | 2/32<br>**6.3** | 2/32<br>**6.3** | 3/32<br>**9.4** | 4/32<br>**12.5** | 2/32<br>**6.3** | 3/32<br>**9.4** | 3/32<br>**9.4** | 7/32<br>**21.9** | |
| Gramática | | | | | | | | 10/34<br>**29.4** | |
| Lectura | | 2/34<br>**5.9** | | 1/34<br>**2.9** | 1/34<br>**2.9** | 4/34<br>**11.8** | 1/34<br>**2.9** | 9/34<br>**26.5** | 9/34<br>**26.5** |
| Total | 2/100<br>**2** | 4/100<br>**4** | 3/100<br>**3** | 5/100<br>**5** | 3/100<br>**3** | 7/100<br>**7** | 4/100<br>**4** | 26/100<br>**26** | 9/100<br>**9** |

FR: frecuencia relativa con N=6 jueces para Comprensión auditiva y Lectura y N=4 para Gramática.
Las celdas en blanco representan consenso o 75% de acuerdo de los jueces respecto a la respuesta 1 ("De acuerdo)" con la aseveración que define cada aspecto evaluado.

## 7.2. Validez de constructo.

La validez de una prueba se refiere a las evidencias que demuestran que la prueba mide el atributo que debería medir de acuerdo con su propósito y su diseño. Por lo tanto es posible plantear que los hallazgos de esta investigación aportan las siguientes evidencias de que el EXEDII evalúa el constructo definido como "manejo del inglés como lengua extranjera al nivel intermedio en los alumnos egresados de la UABC".

1. Con respecto a los Nodos de los reactivos del EXEDII tal como se explica en el apartado 7.1, la relevancia, la representatividad y la pertinencia de los enunciados que describen la tarea de los reactivos fueron valorados favorablemente por los expertos, en la mayoría de los reactivos. No es aventurado interpretar que un juicio favorable en ese sentido representa una evidencia de validez de constructo, toda vez que los nodos son las definiciones de las tareas del dominio que debería medir el EXEDII.

2. A propósito de las características de los reactivos, comentadas en el inciso 2 del apartado de Confiabilidad, las opiniones de los jueces también aportan evidencias de validez de constructo al encontrar similitud entre las tareas del reactivo y las tareas que en un contexto natural encontraría el sustentante. Esta valoración ayuda a responder positivamente a la pregunta de investigación sobre si los reactivos del EXEDII constituyen una muestra del DULM. Por otra parte, el nivel apropiado es una de las características más importantes del reactivo, tanto que se incluye en la definición del constructo, por lo que la evaluación positiva de los expertos en este rubro también puede considerarse como una aportación a la validez del constructo. En los rubros relativos a la adecuación de las características de los reactivos en relación con los sustentantes, los jueces evalúan este aspecto de manera más favorable que el que establece que las características de los reactivos reflejan su contexto cultural. Este último obtuvo observaciones en el sentido de no representar el contexto cultural de los sustentantes en un 21.8% en el caso de la subescala de Comprensión auditiva, el 29.4 en la de Gramática, siendo éste el único aspecto calificado desfavorablemente en esta subescala y el 26.5 en la subescala de Lectura. Es necesario mencionar también que en la subescala de Lectura, el 26% de los reactivos fueron evaluados con señalamientos de provocar sesgo, lo cual llama la atención porque constituye una crítica que cuestiona la validez de constructo y de contenido. Entonces, si las tareas de la prueba son similares a las del mundo real, son del nivel apropiado y reflejan las características de los sustentantes puede plantearse la pregunta ¿es suficiente esa evidencia para determinar que las tareas del EXEDII representan una muestra del Uso del lenguaje? La respuesta no es sencilla, considerando que una de las características de los reactivos no obtuvo valoraciones suficientemente contundentes por parte de los jueces. Más adelante se plantea nuevamente esta pregunta, con el análisis de otros hallazgos de la investigación que puedan ayudar a obtener una respuesta.

3. De acuerdo con las opiniones de los jueces que evaluaron las competencias que mide el EXEDII comparándolas con el DULM, en la Tabla 7.2 se observa que todas las competencias con excepción de dos: (la 4 que corresponde a "comprender a hablantes nativos de una variedad conocida, si hablan

despacio" y la 14 que corresponde al uso de la forma posesiva: 's") son consideradas por los expertos como competencias evaluadas por los reactivos del EXEDII. Esta información constituye al mismo tiempo una evidencia de validez de contenido y de constructo del examen como instrumento (con excepción de algunos reactivos que se analizan más adelante) ya que se encuentra que el contenido que miden los reactivos corresponde con el dominio que debería medir la prueba, a juzgar por las opiniones de los jueces.

Las evidencias de validez de contenido, particularmente cuando se trata de un examen de certificación constituyen también evidencias de validez de constructo (Messick, 1993). Más aún, los exámenes de certificación por el propósito que persiguen deben estar orientados a un criterio (Nitko, 1994, Popham, 1990) y la alineación del criterio con el propósito de la prueba es una evidencia de validez de constructo.

Como se mencionó en el apartado 2.3 de este trabajo de tesis los exámenes criteriales como el EXEDII son construidos con una metodología que obliga a validar los diferentes aspectos del examen como su diseño, la selección de los contenidos, las especificaciones de las tareas, las características de los reactivos, la rúbrica y la administración del examen. Por otra parte, de acuerdo con Bachman y Palmer (1996) las tareas del DULM tienen características que las definen como miembros de ese dominio por lo que, para construir o evaluar una prueba del manejo de la lengua es conveniente verificar que las tareas de la prueba sean esencialmente iguales a las que realizaría el sustentante en un ambiente distinto del de la prueba (Ver apartado 4.5.3 de esta tesis). De acuerdo con esos autores es crucial asegurar que las características de la administración, de la rúbrica, del estímulo, de la respuesta esperada y de la relación entre estímulo y respuesta sean las adecuadas, si se pretende ofrecer evidencias de validez de un instrumento de medición del manejo de un idioma. Si bien en este trabajo de tesis no se hizo un desglose de las características de las tareas del DULM en los aspectos que Bachman y Palmer proponen, sí es posible afirmar que se siguieron todas las recomendaciones de autores que proponen métodos rigurosos para la construcción de instrumentos como el EXEDII (Ver apartado 3.1 de esta tesis) y que el método utilizado es similar, aunque menos específico en los detalles. Dado que en este apartado se están analizando todas las evidencias de constructo que proporciona la investigación aquí reportada, es posible afirmar que se cuenta con evidencias de validez de contenido provenientes de la rigurosidad de la metodología utilizada en la construcción del EXEDII y que estas evidencias aportan elementos para la validez del constructo. En el siguiente apartado se presentan las evidencias provenientes de la indagación de la calidad psicométrica del EXEDII, como instrumento de medición, como una evidencia de validez de constructo.

Tabla 7.2. Porcentaje de reactivos que miden las Competencias/habilidades que están representadas en el DULM.

| Subes-calas | COMPETENCIAS REPRESENTADAS EN EL DULM PARA COMPRENSION AUDITIVA, GRAMATICA Y LECTURA | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 |
| Comp. Aud | 1 | 4 | 1 | 0 | 8 | 6 | 6 | 1 | 1 | 4 | 5 | 14 | | | | | | | | | | | | | | | | | | | | | | | | | |
| Gramática | | | | | | | | | | | | | 34 | 0 | 20 | 3 | 5 | 6 | 4 | 9 | 2 | 1 | 1 | 2 | 22 | 23 | 14 | | | | | | | | | | |
| Lectura | | | | | | | | | | | | | | | | | | | | | | | | | | | | 34 | 34 | 34 | 34 | 34 | 34 | 34 | 34 | 34 | 34 |
| %Total de reactivos que miden | 1 | 4 | 1 | 0 | 8 | 6 | 6 | 1 | 1 | 4 | 5 | 14 | 34 | 0 | 20 | 3 | 5 | 6 | 4 | 9 | 2 | 1 | 1 | 2 | 22 | 23 | 14 | 34 | 34 | 34 | 34 | 34 | 34 | 34 | 34 | 34 | 34 |

Las competencias están numeradas del 1 al 37 y en la tabla inferior se ofrecen las descripciones de las competencias/habilidades. N= 100 reactivos.

| Competencias/habilidades incluidas en el DULM | |
|---|---|
| 1. Discriminar registro formal/informal. | 19. Pronombres personales. |
| 2. Discriminar opiniones de interlocutores | 20. Preposiciones |
| 3. Comprender preguntas de un interlocutor | 21. Conjunciones. |
| 4. Comprender a hablantes nativos de una variedad conocida. | 22. Oraciones negativas. |
| 5. Comprender/expresar inconvenientes | 23. Palabras interrogativas |
| 6. Comprender instrucciones y tareas en clase. | 24. Verbos regulares e irregulares |
| 7. Comprender diferentes puntos de vista en pasado, presente y futuro. | 25. Tiempos: presente y pasado simple, continuo, perfecto y futuro simple. |
| 8. Comprende la solicitud de información de sí mismo. | 26. Verbos auxiliares |
| 9. Comprende solicitud/proporciona ayuda en asuntos simples y dentro de área de conocimiento | 27. Verbos modales. |
| 10. Comprende el intercambio de opiniones y formulación de preguntas con sus pares. | 28. Comprender textos con información de rutina en artículos de divulgación. |
| 11. Comprende conversaciones de temas abstractos o culturales conocidos. | 30. Comprender textos con información no rutinaria ede temas conocidos. |
| 12. Comprende la fundamentación de opiniones | 31. Comprender cartas o notas sobre temas conocidos o predecibles. |
| 13. Uso de sustantivos propios singular/plural | 32. Comprender instrucciones o mensajes básicos como catálogos computarizados de bibliotecas. |
| 14. Forma posesiva. | 33. Intercambiar notas y solicitudes cortas a colegas o conocidos en temas de su trabajo. |
| 15. Artículos demostrativos. | 34. Comprender textos sencillos en periódicos, cartas rutinarias en hoteles y servicios en general. |
| 16. Adjetivos. | 35. Comprender textos sobre opiniones personales. |
| 17. Comparativos. | 36. Comprender ensayos/artículos/reportes de temas técnicos/académicos en área de estudio/trabajo. |
| 18. Superlativos. | 37. Tomar/leer notas en clase apoyándose en lengua materna. |

El dato en las celdas blancas indica el número de reactivos que miden la competencia en la subescala (filas 1, 2 y 3) y el porcentaje total de reactivos que la miden en todo el examen (fila 4).

4. La primera pregunta de investigación de esta tesis plantea la posibilidad de que el EXEDII constituya una escala de reactivos que midan conjuntamente un constructo. Por ello el primer objetivo de esta tesis establece la necesidad de analizar la dimensionalidad del examen. En el inciso 3 del apartado 7.1 se ofrecen los resultados del Análisis de Rasch que apuntan hacia la confiabilidad del EXEDII. Conviene recordar aquí que la confiabilidad de un examen es un requisito indispensable para la validez del mismo (Anastasi, 1977). En ese sentido los resultados obtenidos a propósito de la confiabilidad aportan evidencias de validez de constructo. Pero el hecho de que todos los reactivos de una prueba midan consistentemente un constructo no es suficiente evidencia de que se está midiendo el constructo adecuado. Se requieren otro tipo de evidencias que indiquen que el constructo medido tiene las características adecuadas, de acuerdo con el propósito del examen y de la teoría en la que se sustenta.

Dado que la definición del constructo del EXEDII establece el nivel de las habilidades y/o competencias evaluadas, los hallazgos en este sentido pueden aportar evidencias de validez de constructo. El Análisis de Rasch arroja una calibración de los reactivos en promedio de .00 lógitos con una desviación estándar de .90, cubriendo un rango de -2.01 a +1.80 lógitos. Estos datos indican que los reactivos del examen abarcan un rango amplio de dificultad, lo que es adecuado para un examen de certificación, pero la mayoría de ellos se ubican en un valor en lógitos intermedio, en el que la probabilidad de una respuesta correcta al reactivo es de .50 dada una habilidad intermedia. En la Figura 6.1 del capítulo 6 de esta tesis se presenta el mapa que relaciona a la dificultad de los reactivos con la habilidad de los sustentantes y se observa que la distribución de las medidas de los sustentantes y la calibración de los reactivos se asemeja a la curva normal, y que la mayor parte de los reactivos se asocian a sustentantes con medidas de habilidad alrededor de 0, o habilidad media disminuyendo hacia -1 y aumentando hacia +1. Paulatinamente la medida de los sustentantes se aleja de 0 en ambas direcciones y se observan algunos sustentantes con medidas de habilidad cercanas a 5 lógitos en el extremo positivo y -3 lógitos en el extremo negativo. No obstante la mayor parte de los sustentantes y de los reactivos se localizan alrededor de los 0 lógitos en un rango de -1.5 a +1.5. Entonces, los hallazgos obtenidos a partir del Análisis de Rasch proporcionan evidencias de que el EXEDII constituye una escala homogénea que mide un constructo cognoscitivo de dificultad intermedia, con reactivos que pueden ser contestados primordialmente por sustentantes con medida de habilidad intermedia, pero que sustentantes con habilidad superior pueden contestar los reactivos más difíciles y aquellos sustentantes con medida de habilidad baja responden acertadamente sólo a los reactivos con calibración baja. Estos resultados pueden ser interpretados como evidencia de validez de constructo en el sentido de que las características psicométricas del instrumento coinciden con las del nivel de medición para el que fue diseñado.

Otra información proporcionada por el Análisis de Rasch se observa también en la mencionada Figura 6.1 (mapa reactivos/sustentantes). Conviene recordar que los reactivos que corresponden a la subescala de Comprensión auditiva son del 1 al 32, los de Gramática del 33 al 66 y los de Lectura del 67 al 100. En la figura se han escrito en itálicas, con negritas y en un tipo de letra distinto los reactivos de las subescalas, respectivamente para poder diferenciarlos visualmente. Así, es posible observar que si bien los reactivos de las tres subescalas se agrupan mayoritariamente hacia el valor de 0 lógitos, también hay reactivos en los extremos de ese mismo agrupamiento. Los reactivos que se alejan más hacia los valores más altos son los reactivos 76, 89, 91 y 100, mientras que lo que se alejan en sentido inverso son el 9, 10 y 42, así como 41, 45,46, 47 y 85. En el primer caso los reactivos corresponden a la subescala de Lectura y en el segundo a las subescalas de Comprensión auditiva y de Gramática, salvo un reactivo (85) que corresponde a Lectura. En las tablas 3.1, 3.2 y 3.3 del capítulo 3 de esta tesis se puede observar que los reactivos 76, 89, 91 y 100, todos de la subescala de Lectura corresponden a las tareas de complejidad mediana y alta, atendiendo a su diseño ya que implican habilidades de comprensión de información en párrafos, o de razonar a partir de la información proporcionada en palabras o enunciados. En el caso de la complejidad de los reactivos 9, 10 y 42 se observa que de acuerdo con su diseño estos reactivos estarían midiendo habilidades como identificar palabras o información específica en un diálogo, o el uso de adverbios de una palabra, lo cual no representa teóricamente una dificultad mayor. Si se considera que desde su construcción los reactivos más difíciles del EXEDII corresponden a la subescala de Lectura y que en un estudio anterior que analizó los reactivos de ese examen fueron los reactivos de Lectura los que resultaron con un valor **"p"** más alto (Larrazolo y Velasco, 2000a; Velasco, Larrazolo, Antillón y Rosas, 2005) entonces el hallazgo anterior adquiere relevancia y puede considerarse una evidencia de validez de constructo.

Es claro que no se tiene la expectativa de que todos los reactivos guarden coherencia entre su diseño y su comportamiento en la realidad, pero los hallazgos referidos en el párrafo anterior pueden interpretarse como una evidencia de validez de constructo puesto que en los casos mencionados, que representan los comportamientos extremos se observa congruencia entre el diseño de los reactivos, basado en la teoría sustantiva y el comportamiento real de los reactivos del EXEDII. Los reactivos que tienen una calibración de 0 lógitos, correspondiendo con los sustentantes de medida de habilidad mediana son reactivos cuyo diseño implica tareas de mediana complejidad.

El Análisis de Rasch realizado proporciona otra evidencia de validez de constructo, derivada de la información proporcionada en la Tabla 6.3, en la que se presenta la estructura gráfica del examen. En esta

tabla y en las figuras 6.2 y 6.3 se observa la coherencia entre el examen y la muestra de sustentantes. Conviene recordar que este análisis trabaja elaborando primeramente un modelo contra el cual posteriormente se comparan los datos reales. Las curvas dicotómicas en las figuras mencionadas permiten observar que la expectativa del modelo coincide con el comportamiento de la muestra analizada, con excepción de un desajuste menor en la medida de -3 lógitos. Por ello la estructura del examen se puede considerar como una evidencia de validez de constructo.

5. La segunda pregunta de investigación plantea la posibilidad de que la estructura factorial del EXEDII sea congruente con el diseño de la prueba. Para responderla, el inciso 2 del primer objetivo de esta tesis propone indagar si las relaciones entre las variables del examen pueden ser explicadas por un número reducido de variables latentes de manera congruente con el diseño del examen. El Análisis Factorial Exploratorio es adecuado para investigar patrones de convergencia entre diferentes medidas. La varianza de cada variable observada se explica en términos de cargas factoriales en un número de factores, que puede conocerse de antemano, o solamente suponerse. Derivado de la pregunta y del objetivo se decidió efectuar un análisis factorial exploratorio debido a que desde el punto de vista teórico se esperaría una estructura factorial de tres factores, pero desde el punto de vista empírico era necesario primero realizar un análisis sin restricciones.

La decisión aparentemente fue correcta porque la solución factorial obtenida arroja tres factores, pero el primero de ellos contiene a un mayor número de reactivos que los otros dos, lo que podría significar que existe una escala que mide una habilidad general y dos escalas secundarias que estarían midiendo componentes de esa habilidad. Este resultado es congruente con un estudio similar realizado por Bachman, Davidson, Ryan y Choi (1995) en el que hicieron un AF a los resultados de dos exámenes de certificación del inglés como lengua extranjera y encontraron un factor general y otros factores secundarios. Esos autores denominaron al factor general "Factor de manejo general del lenguaje" (*general language proficiency factor*). Otros autores también han hablado de una habilidad general de manejo del lenguaje (Widdowson, 1996). Esta suposición resulta plausible si se considera la complejidad del lenguaje humano.

Suponer que existe una o más variables que agrupan a todas las variables observadas, reduciéndolas a unas cuantas es posible si se comprende que el procedimiento estadístico que se eligió es capaz de indicar la fuerza y la dirección de las interrelaciones entre las variables, a través de un coeficiente de correlación (Bachman, 2004). El AFE trabaja sobre la matriz de correlaciones buscando la solución que mejor representa las relaciones entre las variables y en la solución factorial encontrada en esta

investigación existen elementos para concluir que se trata de un factor de habilidad general y dos componentes de esa habilidad.

Con base en lo anterior, las cargas factoriales obtenidas en esta investigación aportan evidencia de validez de constructo. En la Tabla 6.4 del capítulo 6 se presentan los factores obtenidos, con sus respectivas cargas factoriales. Primeramente se encuentran tres factores conformados por reactivos que cargan directamente en un número considerable de casos y además, estos reactivos muestran afinidad con el contenido conceptual de su diseño, lo que es un indicio de que el conjunto de datos analizados posee estructura simple. Cuando se habla de Estructura simple se hace mención de un principio de Thurstone (1947, citado en Thompson, 2005, p. 81) según el cual se busca un *patrón teórico de la dispersión de los patrones de los coeficientes* cercanos a *cero y lejanos de cero* que optimizan la interpretabilidad de los factores. Para encontrar estructura simple, la matriz de cargas factoriales debe reunir tres características: a) cada factor debe tener pocas cargas altas y otras cercanas a cero; b) cada variable debe estar saturada solamente en un factor; c) no deben existir factores con la misma distribución. De esta manera, cada factor debe tener una alta correlación con un grupo de variables y baja con el resto de las variables. Este principio es útil a la hora de examinar las características de las variables que se agrupan en un factor, para encontrar rasgos comunes entre ellas y poder identificar y nombrar al factor. Este hallazgo viene a reforzar el resultado obtenido en el análisis de Rasch, que demostró la dimensionalidad del instrumento.

En cuanto a los factores resultantes, se observa que todos los reactivos correspondientes a la subescala de Comprensión auditiva presentan cargas interpretables en el factor 1, excepto los reactivos 9, 12, 14 y 16. También cargan en este factor ocho reactivos de la subescala de Gramática y quince de Lectura. El Factor 2 agrupa a todos los reactivos de la subescala de Gramática, con excepción de dos que no presentan cargas interpretables.18 de 34 reactivos de esta subescala presentan cargas altas en este factor y bajas en los demás; seis reactivos cargan en ese factor con la carga más alta, aunque en otro factor también presentan una carga menor. Cuatro reactivos de esa subescala presentan cargas superiores a .30 en los tres factores. El tercer factor agrupa 29 reactivos de la subescala de Lectura; diez de ellos lo hacen "limpiamente" y ocho presentan la mayor carga en ese factor, pero también cargan en otro, por lo que se considera que contribuyen a este factor.

La teoría sustantiva de esta investigación sostiene que la habilidad en la lengua (meta) está compuesta por dos componentes: el conocimiento de la lengua y la competencia estratégica (Ver capítulo 4 de esta tesis). El primero a su vez está compuesto por dos diferentes habilidades: el conocimiento

organizacional y el conocimiento pragmático. El primero controla la estructura formal tanto para producir, como para comprender el lenguaje. El segundo permite crear o interpretar discurso, relacionándolo con las intenciones y características del contexto (Bachman y Palmer, 1996, p. 67-69). En este sentido es notable que algunos reactivos del EXEDII se agruparan en el primer factor, a pesar de que por su diseño pertenecen a subescalas diferentes. Este hallazgo apoyaría la suposición de una habilidad general que permite al sustentante comprender la tarea a la que se enfrenta y responder de manera congruente, independientemente de que la tarea implique la comprensión auditiva o lectora, o el reconocimiento del correcto uso de reglas gramaticales. Por otra parte, la habilidad para detectar la correcta organización del estímulo aural o visual, de acuerdo con la estructura gramatical del idioma permite al sustentante del EXEDII reconocer la naturaleza de la tarea. Y en última instancia la respuesta al reactivo incluye a otros componentes del uso del lenguaje en una situación determinada, como son el conocimiento del tópico, las características personales y el esquema afectivo del sustentante, quien apela a la competencia estratégica para decidir cómo responder a la tarea.

Es pertinente recordar aquí lo planteado en el apartado 3.1 del capítulo 3 de esta tesis, en donde se explica el desarrollo del EXEDII y se aclara que el examen fue concebido y construido desde el enfoque comunicativo y fue alineado a los contenidos del curso *On Target* de Purpura y Pinkley (1991), que en ese momento era el currículo que estaba vigente en la Escuela de Idiomas de la UABC. También es oprtuno manifestar que en ese momento se decidió que el examen constase de cien reactivos, repartidos en tres "áreas" que son las subescalas del EXEDII. Los reactivos fueron diseñados para medir la habilidad de comprensión auditiva, la habilidad de comprensión de lectura y la habilidad para la aplicación de reglas gramaticales. El sustento teórico en el que se basó la construcción del EXEDII consideraba que existen esas habilidades, aunque la separación de las mismas es meramente conceptual.

Ahora bien, la teoría en la que se sustenta la investigación de las evidencias de validez del EXEDII reportada en esta tesis ofrece un panorama más amplio, en el que se trasciende la concepción de habilidades receptivas y productivas y se propone un marco conceptual más amplio para construir y evaluar exámenes de idiomas, pero sigue siendo compatible con la teoría que sustenta la construcción del EXEDII, si se considera que los reactivos de este examen estarían evaluando solamente la parte receptiva del manejo del lenguaje. No obstante, los resultados de esta investigación ponen de manifiesto la pertinencia de un marco teórico más amplio, en el que las habilidades no tengan que segmentarse artificialmente, sino que se consideren como habilidades más complejas, que puedan dar cuenta de la

parte productiva y de la receptiva de los usos del lenguaje, aunque los reactivos se diseñen específicamente para la evaluación de un segmento del episodio de uso del lenguaje.

A continuación se muestra la Tabla 7.3 en la que los tres factores han sido nombrados a partir del contenido conceptual de los reactivos que presentan carga interpretable.

Tabla 7.3 Estructura del EXEDII después de eliminar los reactivos que no cargan adecuadamente.

| F1 Manejo general del lenguaje | F2 Organización del lenguaje | F3 Manejo textual del lenguaje |
|---|---|---|
| 1 | 33 | 80 |
| 2 | 35 | 82 |
| 3 | 36 | 84 |
| 4 | 37 | 85 |
| 5 | 38 | 86 |
| 6 | 39 | 88 |
| 7 | 41 | 89 |
| 8 | 47 | 90 |
| 10 | 48 | 91 |
| 11 | 49 | 92 |
| 13 | 50 | 93 |
| 15 | 51 | 94 |
| 17 | 52 | 95 |
| 18 | 53 | 96 |
| 19 | 54 | 97 |
| 20 | 55 | 98 |
| 21 | 56 | 99 |
| 22 | 57 | 100 |
| 23 | 58 | |
| 24 | 59 | |
| 25 | 60 | |
| 26 | 61 | |
| 27 | 63 | |
| 28 | 64 | |
| 29 | 65 | |
| 30 | 66 | |
| 31 | | |
| 32 | | |
| 40 | | |
| 42 | | |
| 43 | | |
| 45 | | |
| 46 | | |
| 47 | | |
| 48 | | |
| 49 | | |
| 68 | | |
| 72 | | |
| 73 | | |
| 74 | | |
| 75 | | |
| 83 | | |

En la tabla anterior se han eliminado los reactivos que contribuyen mínimamente a los factores, y se han dejado los que presentan cargas interpretables y además, comparten contenido conceptual. Esta nueva disposición de reactivos es tentativa y en estudios posteriores se deberá verificar si esta depuración contribuye a aumentar la varianza explicada.

### 7.3 Autenticidad.

Esta cualidad de las pruebas es intuitivamente tomada en cuenta por cualquier persona que elabore un examen ya que los reactivos son elaborados pensando en la población a la que van dirigidos. Pero aunque la autenticidad es un aspecto fundamental del diseño de los reactivos, o tareas del examen y de la prueba en general (Bachman y Palmer, 1996) a menudo se deja a la intuición un aspecto que debería ser un elemento sistemáticamente planeado en el diseño de la prueba y de los reactivos. La Figura 4.2 (pag. 70) del capítulo 4 de esta tesis esquematiza la correspondencia entre el uso de la lengua en la prueba y en situaciones fuera de la prueba.

De acuerdo con la teoría es indispensable que las características de las tareas del examen y de la situación del uso de la lengua sean esencialmente iguales a las tareas y la situación de la prueba. Así mismo, las características del usuario de la lengua en el mundo real deberán ser esencialmente iguales a las características de los sustentantes incluyendo el conocimiento del tópico, el esquema afectivo y la habilidad en el idioma. Es evidente que ninguna situación de examen puede imitar de manera completamente auténtica las condiciones de la prueba, pero quienes construyen un examen, lo usan o lo evalúan tendrán que tomar en cuenta todos los elementos esenciales de la situación real que pudieran ser determinantes. Por ejemplo, aunque en el mundo real en ocasiones no es posible escuchar los estímulos auditivos más de una vez (como en una película, en una conferencia, etc.), en conversaciones cara a cara con frecuencia lo es. Por eso, en el DULM se incluyó una competencia que establece la comprensión del lenguaje hablado por angloparlantes nativos siempre que éstos hablen despacio, por considerarse una situación con una alta probabilidad de aparecer en el tipo de situaciones que enfrentarían los sustentantes del EXEDII. Sin embargo como los reactivos de la subescala de Comprensión auditiva de ese examen no permiten que se escuche el diálogo más de una vez, no es posible evaluar esa competencia.

Las características del uso de la lengua en el mundo real implican un contexto que puede ser altamente variable, pero es posible tratar de incluir en las tareas de la prueba los elementos más importantes. En el caso del EXEDII los estímulos de la subescala de Comprensión auditiva son conversaciones cortas entre un hombre y una mujer que simulan ser estudiantes universitarios platicando sobre un tema escolar, como la presentación de un proyecto. También incluyen temas como la elección de un restaurant para ir a comer, o la problemática de una pareja joven que busca un departamento económico para vivir. Los temas y los diálogos fueron elegidos por los elaboradores de reactivos siguiendo las Especificaciones de los reactivos, pero aunque en algunos casos el estímulo del reactivo cumple con las especificaciones, algún detalle del estímulo puede causar la impresión de falta de autenticidad. Por ejemplo, la voz de uno de los interlocutores puede escucharse inadecuada para la situación, como la de una persona de edad avanzada hablando de asistir a un concierto de música moderna.

Algunos aspectos que en ocasiones pasan inadvertidos al elaborar reactivos tienen que ver con el contexto sociolingüístico de los sustentantes. En el caso del EXEDII se consideró que una característica general de sus sustentantes es su condición de estudiantes universitarios. No obstante, como la UABC ofrece carreras que pueden ser clasificadas en distintas áreas académicas, los temas de los reactivos son generales para evitar sesgar las preguntas. Por ello los textos de la subescala de Lectura son auténticos y se refieren a temas relativos a eventos de la Naturaleza, a programas de televisión, a situaciones deportivas, etc. Aún así, la aseveración que obtuvo menor número de consensos de los jueces fue la que dice "Refleja el contexto cultural de los sustentantes" lo que plantea la interrogante acerca de cuáles serían las características culturales de los sustentantes del EXEDII, ya que se ha señalado que el no incorporar las características socioculturales de los sustentantes a la hora de construir, traducir o interpretar los puntajes de los sustentantes puede llevar a graves consecuencias para éstos y pueden dar pie a decisiones que afectan negativamente a grandes grupos de personas (Solano, Trumbull y Nelson-Barber (2002).

La autenticidad de los reactivos es la cualidad que proporciona credibilidad y naturalidad a las tareas del examen. Es difícil de evaluar porque normalmente se percibe más fácilmente la ausencia de ésta, que su presencia. El conocimiento del contexto cultural y de las características de los sustentantes son elementos importantes para lograr autenticidad, de manera que la introducción de información relativa a las características del contexto y de los sustentantes puede contribuir a elevar la autenticidad. La autenticidad está íntimamente con otra cualidad de los reactivos: la interactividad. En el siguiente apartado

se discuten las dos cualidades y su relación con el contenido y el constructo del EXEDII para valorar si aportan evidencias de validez en ese sentido.

### 7.4 Interactividad.

En el modelo de Bachman y Palmer la interactividad se define como "...el tipo y grado de involucramiento de las características del sustentante en el logro de una tarea de la prueba" (Bachman y Palmer, 1996, pag. 25). La autenticidad y la interactividad de las pruebas son características importantes de un examen de idiomas porque interactúan con la habilidad (conformada por el conocimiento de la lengua y la competencia estratégica), con el conocimiento del tópico, las estrategias metacognoscitivas y el esquema afectivo del sustentante. La autenticidad y la interactividad de las tareas inducen al sustentante a contestarlas de una manera realista, tal y como las enfrentarían en una situación natural. Juntas pueden disminuir o potenciar la posibilidad de sesgo al interactuar con el esquema afectivo de los sustentantes.

La autenticidad tiene estrecha relación con la relevancia de la tarea en la prueba, lo que afecta a la validez del contenido del examen, impactando así la amplitud con la que se pueden generalizar las interpretaciones de los puntajes hacia todo el dominio. La interactividad por su parte está vinculada con el grado en el que el sustentante se involucra personalmente con la tarea, apelando a su conocimiento del tópico, a su habilidad (conocimiento de la lengua y estrategias metacognoscitivas) y a su esquema afectivo para contestar las tareas de la prueba, lo que apunta a la validez de constructo (Bachman y Palmer (1996).

De acuerdo con lo anterior las evidencias de validez que proporciona la evaluación de la autenticidad y la interactividad de los reactivos del EXEDII son las siguientes: a) los jueces que evaluaron los reactivos del EXEDII lo hicieron comparándolos con un contenido más amplio que los contenidos académicos de un curso. Se construyó ese criterio preguntando las opiniones de expertos en inglés, pero también las opiniones de los expertos en los estudiantes de la UABC. Esas opiniones acotaron las descripciones de las competencias que se consideran típicas en las personas que manejan el inglés como lengua extranjera en situaciones laborales, sociales, escolares y turísticas, de acuerdo con los estándares internacionales. Dentro de ese marco de referencia, los expertos coincidieron en que los nodos u objetivos que describen las tareas del EXEDII son relevantes, representativos y pertinentes, lo que puede ser interpretado como una evidencia de autenticidad de las tareas. Lo mismo puede decirse de los consensos logrados respecto de la similitud de los reactivos con lo que serían situaciones de la vida real en las que el

sustentante tuviera que manejar el inglés. La valoración positiva del nivel de las tareas es un indicador de interactividad, lo mismo que la evaluación del grado en el que los reactivos reflejan las características de los sustentantes (Ver tablas 6.11, 6.12 y 6.13). El único rubro que fue calificado de manera negativa es el que se refiere al grado en el que los reactivos reflejan el contexto cultural de los sustentantes.

Aparentemente lo anterior no es obra de la casualidad, sino que obedece a una característica que los jueces percibieron sistemáticamente. Los comentarios que anotaron en la parte correspondiente de la Escala de validación, pero sobre todo los que hicieron de viva voz durante las sesiones de trabajo indican que, aunque las tareas son parecidas a las tareas del mundo real (auténticas), algunos de los estímulos no involucran de la misma manera a todos los sustentantes. Más aún, cuando los jueces valoraron si las competencias del DULM están representadas en los reactivos del EXEDII resolvieron que prácticamente todas ellas lo están. Entonces la pregunta de investigación que cuestionaba si los reactivos del EXEDII constituyen una muestra del uso del inglés como lengua extranjera, al nivel intermedio puede ser contestada de manera positiva solamente parcialmente, porque existen algunas dudas acerca de su interactividad.

Parte de la falta de autenticidad y de la interactividad deviene de la situación misma de la evaluación. Es del conocimiento público que los sustentantes perciben como estresante y no auténtica casi cualquier situación de evaluación. La valoración exclusiva de la parte receptiva del uso del lenguaje podría estar afectando la autenticidad y la interactividad de las tareas porque el sustentante no interactúa con el examen, excepto eligiendo una opción de respuesta entre otras cuatro. Aunque este no es el espacio para discutir las ventajas y desventajas de los exámenes de opción múltiple, es pertinente mencionar que en la Conceptualización del Dominio de Uso de la Lengua Meta del EXEDII, elaborado como parte de esta investigación (Ver apartado 5.2 del capítulo 5 de esta tesis) se incluyen algunas competencias de producción de lenguaje hablado y escrito, porque forman parte de las tareas que el sustentante encontraría en el mundo real. La Conceptualización es una situación ideal con la que se compararon los reactivos del EXEDII, a pesar de que las habilidades productivas, o mejor dicho la parte productiva de la actividad evaluada no está contemplada en los reactivos de ese examen, pero no por ello se asume que no estarían presentes en las tareas del mundo real.

## 7.5. Impacto.

El impacto del EXEDII sobre los estudiantes que presentan este examen se ha modificado notablemente en el último año. En el capítulo 3 de esta tesis se narraron las circunstancias que originaron la construcción del EXEDII, así como el impacto que tuvo sobre los estudiantes durante casi una década. Se consideraba que el EXEDII era un examen de alto impacto porque las consecuencias de no acreditarlo le impedían al estudiante acceder a los trámites para su titulación.

Es obvio que ningún examen debería tener el propósito de detener a los estudiantes que han cumplido con todos los requisitos que la universidad impone antes de graduarse, pero la racionalidad que motivó la creación del EXEDII no era la de detener a los estudiantes, sino responder a la necesidad que los universitarios percibían de asegurar que los estudiantes de la UABC adquirieran una habilidad que se considera importante y a veces indispensable para aumentar las probabilidades de conseguir empleo, o de desempeñar las distintas funciones que los profesionistas deben cumplir en el contexto laboral. La percepción de esa necesidad no parece haber cambiado, si se considera que los panelistas que participaron en esta investigación, la mayoría de los cuales eran docentes en la UABC emitieron opiniones favorables en ese sentido e incluso ayudaron a determinar el tipo y nivel de habilidades que los estudiantes deberían demostrar en el manejo del inglés como lengua extranjera.

Las opiniones de los expertos mencionados coinciden con los argumentos que da la Secretaría de Educación Pública para apoyar la necesidad de instrumentar medidas que faciliten el aprendizaje de esa lengua desde la educación básica (SEP, 2000), aunque también existen opiniones en sentido contrario, pero con argumentos de índole distinta, como sería la preocupación de algunos autores acerca de que el aprendizaje casi obligatorio del inglés en los países en vías de desarrollo fomenta la dependencia de éstos con respecto a los países desarrollados, como los Estados Unidos de Norteamérica (Macedo, Dendrinos y Gounari, (2003). Pero más allá de los argumentos políticos y en atención al tipo de situaciones que los egresados de la UABC como habitantes de la frontera mexico-estadounidense podrían enfrentar al insertarse al campo laboral, así como porque es una necesidad que puede palparse cotidianamente en la búsqueda de información en Internet, o en las revistas técnicas y científicas, o en la práctica del turismo internacional parece razonable que la UABC, como una institución de educación superior debería promover el aprendizaje del inglés.

Si bien es notorio que la UABC ha hecho esfuerzos por incluir el aprendizaje del inglés como lengua extranjera al impartir cursos en la Facultad de Idiomas e incluirlo en el currículo de algunas carreras, también es cierto que en 2007 se derogó el artículo que reglamentaba la obligatoriedad de la acreditación de la lengua extranjera, quedando solamente como una recomendación para los estudiantes de esta universidad el que hagan esfuerzos particulares para desarrollar las habilidades de manejo del inglés como lengua extranjera.

El impacto de un examen como el EXEDII no se limita al mencionado arriba. Existen otras razones por las que vale la pena atender a las consecuencias que tiene un examen sobre los sustentantes, o poblaciones mayores. El impacto de la evaluación en general y de la evaluación de idiomas en particular puede definirse como "...las diferentes formas en las que el uso de una prueba afecta a una sociedad, a un sistema educativo y a los individuos que se incluyen en éstos" (Bachman y Palmer, 1996, pag. 39). Por eso es importante que las personas involucradas en el uso de los exámenes consideren los sistemas de valores sociales, educacionales e individuales que subyacen a los instrumentos en el mejor de los casos, o que reciben las consecuencias de éstos. En el apartado 4.5.3.4. del capítulo 4 de esta tesis se mencionan los niveles en los que pueden impactar las consecuencias de los exámenes, es decir a un nivel micro afectan aspectos de la vida del sustentante del examen y al nivel macro las consecuencias de los exámenes impactan en el sistema educativo y en la sociedad en general.

Las razones por las que el impacto de las pruebas es uno de los elementos para evaluar la utilidad de las mismas no son menos importantes que las que sustentan a los otros indicadores de utilidad tratados hasta este punto. El impacto que tiene un examen sobre los individuos puede ser determinante. Por ejemplo los exámenes como el *Test of English as a Foreign Language* en formato computarizado y aplicado por Internet (TOEFL iBT; *Educational Testing Service*), o el *International English Language Testing System* (IELTS; *British Council,* IDP: IELTS *Australia* y *University of Cambridge ESOL Examinations,* 2005) tienen consecuencias que pueden ser la diferencia entre ser, o no ser admitido a estudiar un posgrado en una universidad en los países del llamado primer mundo. Aunque el impacto de este tipo de exámenes puede verse como una cuestión personal de algunos individuos, también puede pensarse en las oportunidades que países como México pueden tener para que los ciudadanos mexicanos accedan al entorno del conocimiento de vanguardia, sea que se refiera a tecnología, ciencia, arte o cualquier otro contexto, siendo la variable idioma la limitante entre los dos mundos.

Como se dijo en el capítulo 4 para Messick (1989, 1993) la validez es un concepto unitario y como el modelo de validez de Messick integra una dimensión social, el modelo implica que las consecuencias de las pruebas también están relacionadas con su validez. El impacto es el vínculo entre el uso de las pruebas y la dimensión social. Sin embargo, los estudios de validación de pruebas están generalmente enfocados desde la tradición psicológica de la evaluación (McNamara y Carsten, 2006). Esto último no es particular de Messick, sino de todos los autores contemporáneos de él, porque la evaluación es entendida como el proceso de construir una racionalidad acerca de lo que una prueba requiere medir y posteriormente reunir evidencias que sustenten las interpretaciones que se hacen acerca de los puntajes obtenidos en las pruebas. En la búsqueda de las evidencias de validez, el impacto se ha reducido al fenómeno del *washback* (explicar o definir) que a pesar de su importancia para la enseñanza de los idiomas, tiene poca contundencia sobre la validez de los instrumentos.

De acuerdo con McNamara y Carsten (2006) el concepto de Utilidad, desarrollado en términos prácticos por Bachman y Palmer (1996) vuelve manejable un aspecto teórico de la validez de las pruebas, que había sido relegado por falta de una traducción a términos prácticos, para así poder evaluarlo. La utilidad de las pruebas, como se ha venido viendo en este capítulo proporciona un eje alrededor del cual se articulan las evidencias de validez recabadas.

### 7.6 Viabilidad.

El último indicador de la utilidad de las pruebas es de una naturaleza distinta a las demás cualidades revisadas. La viabilidad de las pruebas (o *practicality* en términos de Bachman y Palmer) no se refiere al uso de las pruebas, sino a su instrumentación.

Durante el proceso de planeación y de construcción de los instrumentos de medición educativa la viabilidad juega un papel primordial, porque muchas de las decisiones se toman en función de esta cualidad. Tomando por ejemplo el proceso de planeación y construcción del EXEDII, el tamaño del examen se limitó a la evaluación de las llamadas habilidades receptivas, por razones de viabilidad. Se pensó en la conveniencia de medir otras habilidades, como escritura o conversación, pero la rúbrica se hacía mucho más complicada. Los recursos económicos y de equipo no permitían trabajar sobre esas áreas y se decidió que el hecho de que el sistema con el que se construyó (Sicodex) permitiría la agregación de nuevas versiones y componentes, se iniciaría con las tres subescalas a las que se ha hecho mención en esta tesis.

Bachman y Palmer afirman que una prueba es viable (o práctica) si su diseño, desarrollo y uso no exceden los recursos que se tienen disponibles (pag. 36). Los recursos para instrumentar una prueba pueden ser de tres tipos: **a)** humanos, como son los escritores de reactivos, personal que aplica o califica los exámenes, el apoyo técnico, etc.; **b)** materiales, incluyen los espacios (salón para aplicar el examen) el equipo (computadoras, grabadoras, lector óptico, etc.;) o material (lápices, material bibliográfico, papelería, etc.). **c)** tiempo, tanto para la construcción como para la aplicación de la prueba (de corta o larga duración), la calificación y la entrega de los reportes de resultados. También es necesario considerar los gastos que implican estos recursos, tales como honorarios, salarios y costo de mantenimiento.

En el caso del EXEDII, puede afirmarse que se trata de un examen viable puesto que los recursos que se requerían para su construcción fueron suficientes. Se adquirieron algunas computadoras y monitores, diademas con micrófono, pero también se aprovechó el material y espacio que se utiliza para la aplicación de otros exámenes rutinarios. El personal que lo aplica en la Facultad de Idiomas no requiere de entrenamiento especializado más allá de conocimientos en computación y solamente se requiere personal altamente capacitado para su mantenimiento, renovación de versiones, estadísticas, etc. Dado que los sustentantes pagan una cuota por el derecho a presentarlo, la UABC recibe ingresos por su aplicación, pero en esa misma medida está obligada a responder a la comunidad con un examen que cumpla con todos los lineamientos que como instrumento de evaluación educativa debería observar.

Finalmente, no podría concluirse adecuadamente este trabajo de tesis sin mencionar los aspectos relacionados con aquellos reactivos que presentan problemas de algún tipo, y que requieren de una revisión a fondo, en el marco de un análisis de reactivos encaminado a tomar decisiones sobre ellos. En el siguiente apartado se aborda esa parte de la investigación.

**Reactivos que presentan problemas.**

El 35% de los reactivos del EXEDII presentan alguna característica que podría mejorarse para fortalecer la validez de las interpretaciones que se hacen acerca de los puntajes que obtienen sus sustentantes. Las características tienen que ver con los siguientes aspectos: a) no obtuvieron como mínimo el 75% de los acuerdos de los jueces acerca de la relevancia, representatividad, o pertinencia de sus nodos; b) sus características no son las adecuadas; c) no miden el contenido que deberían de medir; d) no se ajustaron al modelo Rasch e) no cargaron en ningún factor. En la Tabla 7.3.(no corresponde a la

numeración anterior) se presentan los reactivos que pueden mejorarse, así como el o los aspectos en donde se detectaron problemas.

Tabla 7.3. Reactivos del EXEDII que pueden ser mejorados para aumentar las evidencias de validez del examen.

| Reac-tivos | Factores | | | INFIT | | OUTFIT | | Aspectos evaluados | | | Calibra-ción en lógitos | NODOS correspondientes a los reactivos |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | F1 | F2 | F3 | MNSQ 1 a 1.3 | ZSTD -1 a +2 | MNSQ 1 a 1.3 | ZSTD -1 a +2 | Nodos | Caracte-rísticas | Compe-tencias | | |
| *16* | .60 | .060 | .092 | | 2.1 | 1.75 | 2.2 | | | | -0.75 | CA: Identificar información específica: instrumento (con qué se hizo algo) |
| 62 | .160 | .203 | .079 | 1.31 | 2.1 | 1.45 | 2.4 | | | <> | 0.92 | G: Pronombres relativos: como objetos animados directos |
| 76 | .063 | .153 | .190 | 1.34 | 2.6 | 1.41 | 2.3 | | | <> | 0.54 | L: Inferencia del significado de vocabulario a partir del contexto. Clave: marcador que indica comparación, similitud o paralelismo |
| **52** | * | .380 | * | 1.34 | 2.5 | 1.45 | 2.5 | | | <> | 0.59 | G: Pasado simple: Formulación de preguntas |
| **59** | .500 | .538 | .332 | | -2.2 | | -2.2 | | | | 0.23 | G: Forma superlativa de adjetivos cortos y largos |
| *30* | .724 | * | * | | -2.1 | | | | | | 0.13 | CA: Distinguir hecho de opinión: Hecho (al nivel del discurso) |
| *12* | .284 | .060 | .138 | | | 1.40 | | | | <> | 1.21 | CA: Inferir: Toma de decisión a partir de una pregunta de información |
| **34** | .131 | .298 | .090 | | | 1.33 | | | | | 0.91 | G: Cláusulas con *When, While* y *As*: contraste de tiempo Pasado y Pasado progresivo |
| *9* | .140 | .430 | -.003 | | | 2.42 | | | | | -2.01 | CA: Obtener significado con base en el contexto: Identificar lugar con base en lo dicho y en los sonidos ambientales |
| **41** | .340 | * | * | | | 1.38 | | | | | -1.57 | G: Presente Progresivo: Contraste con Presente |
| 70 | .201 | .182 | .243 | | | 1.32 | | | | <> | -0.35 | L: Inferencia del significado de vocabulario a partir del contexto. Clave en el contexto: estructura retórica: ejemplo. Presencia en el texto de ejemplos de lo expresado en la frase en la que se encuentra la palabra |
| 79 | .235 | .122 | .235 | | | 1.39 | | | | | 1.68 | L: Identificación y comprensión de idea principal no textual, paráfrasis de varias oraciones |
| 86 | .244 | .066 | .235 | | | 1.43 | | <> | <> | <> | 1.81 | L: Inferencia del significado de vocabulario a partir del contexto Clave: expansión que explica o ejemplifica el significado de la palabra |
| 69 | .234 | .211 | .132 | | | | | | | | 0.16 | L: Identificación y diferenciación de opinión y hecho: opinión del autor o de terceros |
| *14* | .273 | .220 | .124 | | | | | | <> | <> | -0.51 | CA: Obtener significado con base en el contexto: Inferir fecha a partir del contexto de la oración |
| 67 | .291 | .242 | .282 | | | | | | | <> | -0.18 | L: Identificación y comprensión de idea principal textual |
| 71 | .292 | .167 | .195 | | | | | | | <> | 0.21 | L: Inferencia del significado de una palabra a partir del contexto. Verbo. Clave: complemento u objeto directo del verbo |
| 77 | .295 | .291 | .225 | | | | | | | <> | 0.37 | L: Inferencia del significado de vocabulario a partir del contexto: palabra con afijo |
| 78 | .119 | .122 | .235 | | | | | <> | <> | <> | 0.25 | L: Inferencia lógica: información objetiva. |
| 81 | .198 | .182 | .233 | | | | | | <> | | 0.14 | L: Identificación y diferenciación de opinión y hecho: identificar opción diferente |
| *24* | | | | | | | | | <> | <> | -0.71 | CA: Identificar información específica: Reconocer paráfrasis |
| *25* | | | | | | | | | | <> | 1.22 | CA: Identificar idea principal: Síntesis que mejor resume el contenido de un programa |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *26* | | | | | | | | <> | <> | <> | 0.47 | CA: Distinguir hecho de opinión/discriminar: Identificar aseveración no apoyada por elementos en la narración |
| **58** | | | | | | | | | | <> | -0.09 | G: Forma comparativa de adjetivos cortos y largos |
| *18* | | | | | | | | <> | <> | | -.88 | CA: Inferir elemento de una categoria más amplia (tipo de actividad) |
| *17* | | | | | | | | | | | -1.47 | CA: Identificar información especifica: Localizar respuesta a pregunta especifica |
| 85 | | | | | | | | <> | | <> | -1.54 | L: Identificación y comprensión de información especifica: información siguiente a un marcador de discurso que contribuye a su ubicación |
| 89 | | | | | | | | <> | | | -.23 | L: .Identificación y comprensión de idea principal: selección de título |
| 90 | | | | | | | | | | | -.12 | L: .Identificación y comprensión de idea principal: selección de título |
| 96 | | | | | | | | | <> | <> | 1.34 | L: Comprensión de marcador de discurso: resultado o consecuencia |
| *27* | | | | | | | | | <> | | -.30 | CA: Identificar información especifica: Razón para (no) hacer algo |
| *28* | | | | | | | | | <> | | -.44 | CA: Inferir significado: Palabra (adjetivo) que describe reacción emocional a un acontecimiento |
| **36** | | | | | | | | | <> | | -.09 | G: Simple Present: Questions with Be and a Third-Person Plural Subject |
| **48** | | | | | | | | | <> | | -1.01 | G: Simple Past: Irregular Verbs |
| 80 | | | | | | | | | <> | | .75 | L: Identificación y comprensión de información especifica: información en ejemplo o enumeración |

*No presentan carga factorial interpretable.
Las columnas 5 a 8 muestran os valores que sobrepasan el rango de 1 a 1.3 y de -2 a2 para ajuste interno y externo respectivamente, que es el rango utilizado en este trabajo de tesis. Algunos autores recomiendan .7 como el valor mínimo en lugar de 1, en cuyo caso otros reactivos estarían fuera de la expectativa del modelo (Backhoff, 2008, comunicación personal).
<> No se logró consenso, ni al menos 75% de acuerdo entre los jueces por lo que se interpreta que existen problemas en ese aspecto evaluado.
Las celdas en blanco indican la ausencia de problemas.
En la columna 1 se utiliza la misma convención que en el resto de la tesis: reactivos escritos con *itálicas* pertenecen a Comprensión auditiva; en negritas a Gramática y con un tipo de letra distinto significa que el reactivo es de Lectura.
La última columna indica la subescala a la que pertenece el reactivo y el nodo que mide.

Los resultados anteriormente expuestos constituyen una información valiosa para analizar a fondo cada uno de los reactivos y los problemas detectados, a fin de tomar decisiones sobre lo que conviene hacer con cada reactivo. Toda consideración deberá hacerse con base en la teoría sustantiva y con el afán de mejorar el instrumento. Pero esa es una tarea que queda fuera del ámbito de esta investigación y si la UABC requiere seguir utilizando este instrumento deberá realizarse el estudio para el cual esta tesis proporciona el punto de partida.

**Conclusiones**.

El propósito de la investigación que se reporta en esta tesis era el de recabar evidencias de validez del EXEDII. Se diseñaron y operaron dos estrategias para producir información suficiente con la cual responder a las preguntas planteadas al inicio de la investigación.

Los hallazgos obtenidos aportan evidencias suficientes para concluir que las tareas que miden los reactivos del EXEDII son en su mayoría relevantes y representativas de un dominio que fue definido en la Conceptualización del EXEDII. Las opiniones de los jueces consultados proporcionan evidencias suficientes para considerar que las características de los reactivos son esencialmente similares a las tareas que el sustentante encontraría en la vida real, con la salvedad de las características que reflejan el contexto cultural de los sustentantes. Igualmente, las opiniones de los jueces permiten considerar que las competencias lingüísticas que valora el EXEDII coinciden en su mayoría con las de la Conceptualización del DULM. No obstante, ninguna competencia de producción de lenguaje es evaluada por el EXEDII porque en las especificaciones de su diseño no se contemplan reactivos para habilidades como hablar o escribir.

A pesar de que la UABC definió el nivel que debería tener el examen y el currículo al que se alineó, en esta investigación la metodología de la validación de constructo no siguió el procedimiento que habitualmente se sigue, el cual consiste en comparar el contenido de los reactivos contra un criterio, que a menudo es un currículo determinado. Se compararon los nodos de los reactivos con las competencias que se establecen en los estándares

internacionales, que clasifican las competencias lingüísticas que deben demostrar los estudiantes que desean ingresar a las universidades, o a laborar profesionalmente en los países desarrollados. Estas últimas fueron consultadas en sus niveles intermedio bajo e intermedio medio, para después acotarlas y precisarlas de acuerdo con las opiniones de los dos tipos de expertos consultados. Esta diferencia en la metodología se derivó de la necesidad planteada por la teoría sustantiva de un dominio más amplio y complejo, que aportara los elementos necesarios para valorar la autenticidad e interactividad de los reactivos, características que como se ha venido mencionando, son esenciales para la utilidad de la prueba. No obstante, el giro en la metodología complicó el proceso del trabajo de los jueces, a quienes se les dificultó trabajar con un criterio conceptual que no se expresaba en términos de objetivos, sino de competencias lingüísticas. Posiblemente la discrepancia entre los reactivos que presentan problemas por razones de contenido y por razones de calidad psicométrica tenga que ver, al menos parcialmente, con la forma en la que los jueces respondieron a las preguntas de la Escala de validación. Es decir, los jueces tenían la encomienda de determinar por un lado si el nodo del reactivo pertenecía al contenido a medir y por otro, si cada reactivo en particular efectivamente medía el nodo correspondiente. Las respuestas de cada panelista quedaron anotadas en las escalas de validación de cada subescala, pero las preguntas y los comentarios verbales y escritos de los jueces pusieron de manifiesto que se trataba de una tarea difícil. Hay que decir que se hicieron todos los esfuerzos para dejar claro en todos los jueces la naturaleza de la tarea que se les encomendaba, pero la impresión de la autora de esta tesis es que pudieran haber quedado dudas sin resolver.

Los resultados de las dos vertientes de datos obtenidos se organizaron en torno a la Utilidad de la prueba porque la teoría sustantiva de esta investigación propone a este concepto como la cualidad más importante de cualquier prueba educativa. Por ello, puede concluirse que los análisis efectuados indican que:

**a)** es satisfactorio el grado de **confiabilidad** que arrojan los estimados proporcionados por los análisis estadísticos de los patrones de respuesta de los sustentantes, así como por la proporción de acuerdos de los expertos que evaluaron las características de los reactivos, por lo

que se concluye que los puntajes obtenidos por los sustentantes del EXEDII son un indicador confiable del dominio real de éstos.

**b)** el **constructo** evaluado por los reactivos de la prueba corresponde satisfactoriamente con el tipo de tareas y competencias lingüísticas que encontrarían los egresados de la UABC en el mundo real. No obstante, es necesario acotar el alcance de esta afirmación dentro de los límites de los hallazgos de esta investigación puesto que, si bien se obtienen evidencias contundentes de la calidad psicométrica del instrumento, también existen dudas acerca del grado en que algunas características de los reactivos reflejan el contexto sociocultural de los sustentantes. Estas dudas no cuestionan de fondo la validez del contenido del EXEDII, sino que pueden considerarse como un aspecto concreto y determinado en el que se requiere trabajar para mejorar algunos reactivos del examen. Por lo tanto se concluye que las evidencias de validez de constructo aportadas por esta investigación son suficientes para certificar el manejo del inglés como lengua extranjera al nivel intermedia en los estudiantes que egresan de la UABC.

**c)** de acuerdo con las opiniones de los jueces, las características de las tareas de los reactivos del EXEDII presentan un grado de **autenticidad** aceptable, ya que se lograron consensos y acuerdos mayores al 75% en casi todos los aspectos evaluados, con excepción de la característica que reflejaría el contexto cultural de los sustentantes.

**d)** con base en los consensos y en los acuerdos mayores al 75% de las opiniones de los jueces, los reactivos del EXEDII presentan un grado aceptable de **interactividad**, lo que permite que los sustentantes se involucren con la tarea que representan los reactivos y aumenta la posibilidad de que contesten espontáneamente. No obstante se recomienda estudiar las características que disminuyen la autenticidad de los reactivos dado que la interactividad está estrechamente ligada con la autenticidad de la prueba.

**e)** El EXEDII fue concebido y construido como un examen de alto **impacto** debido a las condiciones en las que se aplicó a lo largo de casi una década. Como tal, el examen fue sometido a varios tipos de análisis estadísticos (no reportados en esta tesis) con el fin de verificar la pertinencia de los parámetros de dificultad, discriminación y confiabilidad, de los

cuales se obtuvieron resultados satisfactorios de los indicadores de calidad psicométrica del examen. Así mismo, se realizó una investigación que determinó el punto de corte técnico del EXEDII, dentro del marco de una tesis de maestría. Aunque el punto de corte fue establecido con un método riguroso, durante casi una década se ha utilizado un punto de corte "funcional" que es sensiblemente mas bajo que el estimado técnicamente. Este punto de corte obedece a criterios administrativos, y aún así una proporción elevada de alumnos no aprobaban el EXEDII. A raíz de la derogación del artículo del reglamento escolar que estipulaba la obligatoriedad de la acreditación del dominio de una lengua extranjera, el EXEDII deja de ser un examen de alto impacto y podría ahora utilizarse el punto de corte técnico para aumentar la certeza de la medición que se realiza con este instrumento.

f) la **viabilidad** es y ha sido una de sus características más robustas del EXEDII. Dado que fue construido con el sistema Sicodex, es posible la inserción de nuevas versiones y otro tipo de desarrollos tecnológicos sin operar grandes modificaciones a la estructura del examen. La UABC cuenta con el equipo y material necesario para su administración ya que es el mismo que se requiere para la aplicación del examen de selección para nuevo ingreso (EXHCOBA, Backhoff y Tirado, 1993; Backhoff, 2001) y para la batería de pruebas psicométricas. Los recursos humanos necesarios para su administración son parte del personal de la Facultad de Idiomas de la UABC, que recibe un porcentaje de las utilidades que se obtienen por el EXEDII. Solamente se requiere personal altamente capacitado para la calificación y mantenimiento del examen, debido al tipo de análisis y estudios que se requiere conducir para garantizar su óptimo desempeño.

## Epílogo

La naturaleza de la evaluación del lenguaje es compleja. El lenguaje como objeto de estudio y la evaluación psicológica y educativa como medios para abordarlo han tenido un largo desarrollo, con numerosos exponentes que han arrojado luz sobre los aspectos más intrincados y difíciles de estudiar.

Debido a su complejidad precisamente, algunos teóricos han hecho intentos por despojar al lenguaje de los atributos que lo vuelven casi inalcanzable, pero no ha habido consenso en cuáles son los atributos de los que se puede prescindir para convertirlo en un objeto de estudio determinado, como los que abordan las ciencias naturales. Algunas corrientes de pensamiento han reclamado para sí el complicado asunto de la adquisición del lenguaje humano y las circunstancias en las que se da, comparándolo con el lenguaje de otras especies de animales. El lenguaje ha sido definido como herramienta, como órgano, como producto del pensamiento y como producto de la cultura en la que se adquiere. Pero el lenguaje como objeto de estudio ha logrado escabullirse de los más brillantes pensadores y de las teorías más completas, porque siempre hay elementos que quedan fuera de cualquier aproximación que lo intente explicar.

Si la explicación del lenguaje humano es difícil, su medición es aún más complicada y las diversas teorías han generado metodologías para definirlo y evaluarlo. El enfoque desde el que se le explica, enseña y evalúa ha cambiado y en la actualidad parece haber un acuerdo acerca de que la mejor forma de estudiarlo es mediante un enfoque comunicativo. No obstante, el acuerdo llega hasta allí y a partir de ese punto se diversifican las definiciones de enfoque comunicativo.

En lo que sí hay consenso es en la necesidad de seguir buscando mejorar los métodos para evaluar el manejo del lenguaje para hacerlo de manera más incluyente, más justa y que refleje todos los aspectos que hacen del lenguaje humano una característica definitoria de la especie, aunque como herramienta de comunicación no sea exclusivo de ésta.

Los hallazgos de esta investigación indican que el EXEDII es un instrumento robusto, con una calidad técnica sobresaliente y con un poder psicométrico contundente. No obstante, se perciben fallas menores que pueden subsanarse mediante un plan riguroso, conducido por expertos y realizado por personal adecuado. En el apartado de recomendaciones se plantean los aspectos que requieren ser trabajados.

Quizá podría haberse puesto el punto final a esta investigación en el párrafo anterior, pero hubiera quedado pendiente un tema que permea en todo el trabajo realizado: la dificultad

de medir el lenguaje más allá de los aspectos cognoscitivos. Si el lenguaje es un medio para la comunicación social, debería evaluarse la dimensión social de esa comunicación.

Las tendencias recientes de evaluación del lenguaje han empezado a ampliar el espectro del objeto de medición y se han hecho esfuerzos por incluir todos los elementos que hacen del lenguaje un constructo tan difícil de aprehender. La investigación que aquí se reporta tiene un pie en el piso firme de la medición psicométrica de los aspectos cognoscitivos del lenguaje, y el otro en el suelo resbaladizo de los aspectos sociales del lenguaje. Se decidió combinar la metodología rigurosa y ampliamente probada de la psicometría con una aproximación que señala y define los aspectos que son relevantes, dependiendo del propósito de cada evaluación. El modelo de Bachman y Palmer (1996) fungió como un puente entre el terreno conocido y el que queremos conocer. Por esa razón no se instrumentó un estudio de validez de contenido como normalmente se hace: cotejando los objetivos de los reactivos del examen con un criterio establecido, como por ejemplo un currículo, sino que se creó un criterio "a la medida". Tampoco se investigó la validez del constructo "corriendo" un Análisis Factorial, sino que se indagó primero su dimensionalidad y una vez habiendo establecido que se trata de un instrumento que mide una sola dimensión, se procedió a investigar su estructura. Finalmente, se integró toda la información resultante alrededor del concepto de la Utilidad de la prueba como un eje articulador, cuya importancia se puso de manifiesto a lo largo de la investigación. Partiendo de la teoría de la validez de Messick, y operacionalizando los conceptos con la ayuda de Bachman y Palmer se logró contestar las preguntas y cumplir con los objetivos de la investigación y aunque se lograron todos los propósitos, las dos vertientes de la investigación desembocaron en un océano de preguntas y posibilidades de investigación en las que será inevitable navegar en el bote de la dimensión social del lenguaje.

Recomendaciones

La investigación realizada para recabar evidencias de validez del EXEDII produjo resultados satisfactorios puesto que no solamente se obtuvieron evidencias de las fortalezas del examen, sino que se detectaron puntualmente sus debilidades. A continuación se plantean los aspectos que deberán ser atendidos para seguir utilizando el EXEDII con la confianza de que se está midiendo lo que se pretende medir.

1. El EXEDII fue concebido y construido como un examen que evalúa las "habilidades receptivas", es decir comprensión auditiva y comprensión de la lectura, así como también la habilidad para aplicar las reglas gramaticales que el sustentante conoce. La primera recomendación que se hace se deriva de esa característica fundamental del examen, característica que se definió desde la etapa de su diseño. La recomendación más importante que puede hacerse al EXEDII es la de complementar el dominio que mide con la evaluación de las "habilidades productivas."

2. Para complementar el EXEDII con los elementos que le darían la posibilidad de medir las competencias lingüísticas de hablar y escribir en inglés, sería necesaria la modificación de los reactivos cuyas fallas han quedado documentadas en este trabajo de tesis. En algunas ocasiones será necesario modificar la especificación del reactivo, y en otras solamente aspectos de la estructura del reactivo. Existen diferentes métodos para modificar reactivos, pero se sugiere que se elija uno suficientemente riguroso y apegado a los estándares internacionales.

3. Una aportación que hace este trabajo de tesis es la Conceptualización de lo que debería medir el EXEDII. Es de notar que la conceptualización incluye la valoración de competencias que se miden sólo parcialmente en el EXEDII. Competencias como "comprender el discurso de los interlocutores en una conversación cara a cara" son abordadas en el EXEDII solamente de manera

parcial. Se sugiere abordar la tarea desde el modelo de Bachman y Palmer (1996), o cualquier otro método riguroso.

4. Los jueces que evaluaron las características de los reactivos del EXEDII no llegaron a consensos al respecto de algunos rubros evaluados. Se recomienda modificar algunas características de los reactivos que obtuvieron menos del 75% de acuerdos de los expertos.

5. Con relación a las características del formato del examen, se recogen algunas sugerencias que se ha planteado en repetidas ocasiones, dentro y fuera del contexto de esta investigación. Si bien la interfaz del examen fue diseñada para facilitar las tareas del examen, ha habido comentarios acerca de la necesidad de un diseño más moderno, en el que los sustentantes no requieran operar algunos "botones", lo que al parecer complica innecesariamente la tarea.

6. Por otra parte, ha sido recurrente también la sugerencia de posibilitar que se escuche dos veces el diálogo de la subescala de comprensión auditiva, así como la de mejorar la calidad del audio.

## REFERENCIAS


Abedi, J. (2003). Considering English Language Learners in the No Child Left Behind Act. Reporting Adequate Yearly Progress. CRESST/UCLA.

Alderson, J.C., (1993). Judgments in language testing. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing.* Alexandria, VA, USA: TESOL.
Alvarez-Gayou, J. L. (2005). *Cómo hacer investigación cualitativa.* México: Paidós.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1954, 1966, 1999). *Standards for Educational and Psychological Testing.* Washington, D.C.: American Educational Research Association. Consultado el 24 de noviembre de 2003 en: http://www.apa.org/ppo/issues/phtestandx.html

American Council on the Teaching of Foreign Languages (ACTFL). *Standards for Foreign Language Learning.* Internet: 1° de Junio de 2000 y Febrero 10 de 2000
http://www.actfl.org/htdocs/standards.htm 1° de junio de 2000
http://www.actfl.org/htdocs/pubs/guidelines.htm

Anastasi, A. (1977). *Tests Psicológicos.* 3era edición. México. Aguilar.

AskOxford.com. (2005). Oxford dictionaries. UK: Oxford University Press. Consultado en el sitio Web: http://www.askoxford.com/globalenglish/?view=uk el 10 de febrero de 2005.

Bachman, L.F, (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L. F. (2004). *Statistical Analyses for Language Assessment.* Gran Bretaña: Cambridge University Press.

Bachman, L. F., Davidson, F., Ryan, K. y Choi, I.C. (1995). *An investigation into the comparability of two tests of English as a foreign language: The Cambridge-TOEFL comparability study.* Cambridge: University of Cambridge Local Examinations Syndicate y Cambridge University Press.

Bachman, L. y Palmer, A. (1996). *Language testing in practice.* Oxford University Press.

Backhoff, E. y Tirado, F. (1993). Desarrollo del Examen de Habilidades y Conocimientos Básicos. *Revista de la Educación Superior,* 83, 95-118.

Bloom, B. S. (Ed.). 1956, 1984. *Taxonomy of educational objectives: The classification of educational goals*: Libro I, The cognitive domain. Nueva York: D. McKay Company Inc.

Bloom, B. S. (Ed.), (1981). *Taxonomía de los objetivos de la educación: la clasificación de las metas educacionales. El dominio cognoscitivo.* México: El Ateneo

Bock, R. D., Gibbons, R. y Muraki, E. (1988). Full-Information Item factor Analysis. *Applied Psychological Measurement.* Vol. 12, No. 3. Septiembre, pp. 261-280.

Brumfit, C., y Johnson, K. (1979). The Communicative Approach to Language Teaching (eds.). Oxford: Oxford University Press.

Canale, M. y Swain, M., (1980) Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1: 1-47.

CENEVAL. (2000). *Estándares de calidad para Instrumentos de Evaluación Educativa.* Centro Nacional para la Evaluación de la Educación Superior. México.

Chomsky, N. (1971). *El lenguaje y el entendimiento.* 3ª edición. Barcelona: Editorial Seix Barral.

Contreras Niño, L. A. (2000). *Desarrollo y pilotaje de un Examen de español para la Educación Primaria en Baja California.* Tesis de Maestría no publicada. Universidad Autónoma de Baja California, Ensenada, B.C., México.

DiMatteo, D. (2004). *La diferencia entre EFL y ESL.* Consultado el 4 de agosto de 2006 en el sitio Web: http://www.blueturnip.com/projects/edu/english/efl/esl_versus_efl.php

Durán, R.P. (1986) *Validity and language skills assessment: Non-English background students.* En Howard Wainer y Henry I. Braun (Eds.). Educational Testing Services, NJ: Lawrence Earlbaum Associates Publishers.

Ferrando, P.J. (1996). Evaluación de la unidimensionalidad de los ítems mediante análisis factorial. *Psicothema*, Vol. 8, No. 2, 397-410
(*Scientific Software International*, Lincolnwood, IL).

González Montesinos, M.J. (2004). Defining and Measuring Academic Standards for Higher Education: A Formative Study at the University of Sonora. Tesis doctoral no publicada. Universidad de Arizona, E.U.A.

González Montesinos, M.J. (2008). Análisis de Reactivos a través del Modelo Rasch. Serie 1, Medición y Metodología. En: Recursos de Winsteps/Ministep en Español. Consultado el 11 de marzo de 2008 en el sitio Web: http://www.winsteps.com/recursos

HajiPourNezhad, G. (2003). An approach to the validation of judgments in language testing. *Reportes de la Conferencia del año 2003 de la revista Journal of Applied Language Testing, que apareció en el Vol. 7, No. 2 Jun 2003, pp10-12 Consultado en el sitio Web* JALT Testing & Evaluation SIG Newsletter

Halliday, M.A.K., (1994). *El lenguaje como semiótica social.* México: FCE. Título original: Language as Social Semiotic: The Social Interpretation of Language and Meaning (1978).

Hambleton, R.K. (1988). Criterion-referenced measurement. En Keeves, J.P. (Ed.) *Educational Research, Methodology and Measurement: an International Handbook.* (pp. 277-282). Washington: Pergamon Press.

Heaton, J.B. (1988). *Writing English Language Tests.* Nueva York: Longman Inc

Hernández Rojas, G. (2007). La comprensión y la composición del discurso escrito desde el paradigma histórico-cultural. *Perfiles educativos.* [online]. 2005, vol. 27, no. 107 [citado 2007-10-05], pp. 85-117. Disponible en: <http://scielo.unam.mx/scielo.php?script=sci_arttext&pid=S0185-26982005000000005&lng=es&nrm=iso>. ISSN 0185-2698.

Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, P. (2003). *Metodología de la Investigación.* Tercera edición. México: McGraw Hill

Hilgard, E.R. (1987). *Psychology in America.* EUA: Harcourt Brace Jovanovich.

Holmes, D. y Duron, S. (2000, abril). Introducción. *LEP Students and High-Stakes Assessment.* Consultado en el sitio Web de The George Washington University, Graduate School of Education on Human Development, NCELA: http://www.ncela.hwu.edu/pubs/reports/highstakes/intro.htm

Hopkins, K.D. (1998). *Educational and psychological measurement and evaluation.* (8a. ed). EUA: Allyn y Bacon.

Hughes A. (1989). Testing for language teachers. *Cambridge handbooks for language teachers.* Gran Bretaña: Cambridge University Press.

Hui Ling Tsai, C. (2003). Issues of validity in the assessment of writing performance. *TESOL & Applied Linguistics,* Vol. 4, No. 2 The Forum. Teacher's College, Columbia University working papers.

Hymes, D. (1971). Competence and performance in linguistic theory. En: R. Huxley y E. Ingram (eds.), *Language Acquisition: Models and Methods.* Nueva York: Academic Press, 3-24

Jaeger, R. (1993). Certification of student competence. En R.L. Linn (Ed.) *Educational Measurement.* 3a. ed. American Council on education. Series on Higher Education. EUA: Orix Press.

Kachigan, S. K. (1991). *Multivariate stastistical Analysis: a conceptual introduction.* 2ª ed. Nueva York: Radius Press.

Larrazolo Reyna, N. y Velasco Ariza, V. (2000 a). *Examen de Egreso del Idioma Inglés (EXEDII), Índices de dificultad y discriminación.* Memorias del Cuarto Foro de Evaluación Educativa (pags. 111-115). México.

Linacre, J. M. (2006). *A User's Guide to Winsteps Ministeps: Rasch-Model Computer Programs.* Chicago, IL: Electronic Publication. www.winsteps.com

Loera Varela, A. (2000). *Los grupos de enfoque en la investigación cualitativa.* INDES-BID. Consultado en el Sitio Web:
http://www.google.com.mx/search?hl=&q=grupos+enfoque&meta=

Macedo, D., Dendrinos, B. y Gounari, P. (2003). *Lengua, ideología y poder.* La hegemonía del inglés. Barcelona: Editorial Graó.
Martínez Rizo, F. (2001). Evaluación educativa y pruebas estandarizadas. Elementos para enriquecer el debate. *Revista de la Educación Superior.* Vol. XXX (4), No. 120. Octubre-Diciembre. Pp.71-85.

Martínez Rizo, F. (2004). ¿Aprobar o reprobar? *Revista Mexicana de Investigación Educativa.* Vol. 9. Num. 23. pp. 817-839

McNamara, T. y Carsten, R. (2006). *Language Testing:* The Social Dimension. Reino Unido: Blackwell Publishing, Ltd.

Messick, S. (1988). The once and future issues of validity: assessing the meaning and consequences of measurement. En Wainer, H. y Henry I Braun (Eds.). *Test validity.* NJ:Lawrence Earlbaum Associates Publishers.

Messick, S. (1993). Validity. En R.L. Linn (Ed.). *Educational Measurement.* Phoenix: AmericancCouncil on Education y The Orix Press. American Council on Education. Series on Higher Education. Pp 13-103.

Moreno Bayardo, M.G. s.f. *El desarrollo de habilidades como objetivo educativo. Una aproximación conceptual.* Consultado el 12 de noviembre de 2007 en el sitio Web: http://educar.jalisco.gob.mx/

Migration Policy Institute (2004). Washington, D.C. Consultado en el sitio Web: source@migrationinformation.org el 10 de febrero de 2005.

Nitko, A.J. 1994 (Julio). A Model for Development Curriculum-Driven Criterion-Referenced and Norm-Referenced Examination for Certification and Selection of Students. Artículo presentado en la Conference of Education, Evaluation and Assessment for the Association Studies of Educational Evaluation in Sudafrica (ASEESA). Sudáfrica.

Nunnally, J.C. y Bernstein, I.J. (1995). *Teoría Psicométrica.* Tercera ed. México: McGraw Hill.

Oller, J.W., (1979). *Language tests at school: A pragmatic approach.* Londres: Longman.

Piaget, J. (1984). *El lenguaje y el pensamiento del niño pequeño.* Traducción de Elba Mendolia. Barcelona: Paidós

Popham, W. J. 1990. *Modern Educational Measurement. A practitioner's Perspective.* Segunda edición. Washington: Allyn y Bacon.

Purpura, J. E. y Pinkley, D. (1991). On Target 2. *Scott Foresman English, Intermediate level. 2ª* edición. EUA: Pearson Education

Real Academia Española (2001). *Diccionario de la lengua española. Vigésima segunda edición.* Consultado el 2 de octubre de 2007 en: http://buscon.rae.es/draeI/

Reglamentos Universitarios. (1995). Publicado por la Dirección General de Servicios Escolares de la UABC, Departamento de Diseño, DGEU/UABC.

Santos Caicedo, D. A. (2007). Saussure y Chomsky. *Un análisis paradigmático de los aportes de F. de Saussure y N. Chomsky al campo de los estudios del lenguaje*. Consultado el 26 de marzo de 2008 en el sitio Web: http://www.yontorres.blogspot.com/2007/08/saussure-y-chomsky.html

Secretaría de Educación Pública, (2005). La enseñanza de las lenguas extranjeras en México. El lenguaje y el desarrollo. Autor. Consultado en el sitio Web PortalSEP: http://www.sep.gob.mx/res/sep/sep_1001_05/7720?op=1

Skinner, B. F. (1981). *Conducta Verbal.* México: Trillas

SparkNotes (s.f.). *English Grammar SparkCharts™*. Barnes & Noble. Consultado en el sitio Web:

http://search.barnesandnoble.com/booksearch/isbninquiry.asp?z=y&cds2Pid=6678&isbn=1586636456

Solano Flores, G., Trumbull, E. y Nelson-Barber, S. (2002). Concurrent Development of Dual Language Assessments: an Alternative to Translating Tests for Linguistic Minorities. International Journal of Testing, 2 (2), 107-129.

University of Washington, Office of Educational Assessment. (s.f.). Data Processing, Score Pak, Item Analysis. Consultado el 25 de febrero de 2003 del sitio Web de University of Washington, Office of Educational Assessment: http://www.washington.edu.oea.item.htm

Velasco Ariza, V., Larrazolo Reyna, L., Antillón Macías, L.E. y Rosas Morales, M. (2005). Evaluación del inglés como lengua extranjera: balance de los resultados obtenidos con el Examen de Egreso del Idioma Inglés (EXEDII). Memorias del VII Congreso Mexicano de Investigación Educativa. Hermosillo, Sonora.

U.S. Census Bureau, (2003). Consultado el 26 de junio de 2004 en el sitio Web: http:www.census.gov

U.S. Department of Education (2004). *A Guide to Education and* No Child Left Behind. Consultado en el sitio Web: http://www.ed.gov/nclb/overview/intro/guide/index.html el 10 de febrero de 2005.

Wang, L., Bachman, L. F., Carr, N., Kamei, G., Kim, M., Llosa, L., (2000, marzo). *A cognitive-psychometric approach to construct validation of Web-based language assessment.* Work-in-progress. Documento presentado en el 22 Annual Language Testing Research Colloquium, Vancouver, BC, Canada.

Ward, A.W., Stoker, H.W., Murray-Ward, M. (1996). *Educational Measurement: Origins, Theories and Explications*: (1) Basic concepts and theories. Lanham, MD: University Press of America.

Widdowson, H.G. (1990). *Aspects of Language Teaching,* Oxford, Oxford University Press.

# ANEXO 5.1. Validez de contenido.

## Resultados completos de la valoración de los jueces

En este Anexo se presentan los resultados de la comparación del contenido del EXEDII contra el criterio denominado Dominio de Uso de la Lengua Meta (DULM).

Las tablas que se incluyen aquí fueron elaboradas con los datos obtenidos de los análisis efectuados a las respuestas de los tres paneles de expertos que evaluaron varios aspectos del EXEDII. Los aspectos evaluados se presentan en el Anexo 5.2.

Las respuestas 1 (de acuerdo) ó 2 (en desacuerdo) anotadas en las escalas de validación se vaciaron en una base de datos. El cómputo de los resultados se realizó con ayuda del programa de cómputo SPSS. Se obtuvieron las frecuencias relativas de las respuestas marcadas con 1, las cuales se muestran en las columnas respectivas. Esa proporción se expresa en porcentajes en otra columna.

Los paneles de Comprensión auditiva y de Lectura estuvieron conformados por seis expertos. El panel de Gramática estuvo conformado por cuatro expertos.

Los reactivos de la subescala de Comprensión auditiva son del 1 al 32; los de la subescala de Gramática son del 33 al 66 y los de la subescala de Lectura son del 67 al 100.

La denominación de las tablas incluye la letra A indicando que se trata del presente anexo, seguida del número 5 haciendo referencia al capítulo 5 que trata la indagación de validez de contenido y finalmente un número consecutivo, que indica la tabla correspondiente.

El criterio para considerar que las respuestas de los jueces constituyen una evidencia de validez para los reactivos es de al menos el 75% de acuerdos entre los jueces con respecto del aspecto evaluado y de esa manera se presentan en el capítulo 5. En este anexo se incluyen todas las frecuencias computadas.

Tabla A5.1. Acuerdos entre los panelistas respecto de los aspectos evaluados de los Nodos y Características de los reactivos, expresados en frecuencias relativas y porcentajes.

Subescala de Comprensión auditiva.

| Reactivos | NODOS | | | | | | | | CARACTERISTICAS DE LOS REACTIVOS | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Relevante | | Representativo | | Pertinente | | Claramente descrito | | Similar a la realidad | | Nivel apropiado | | Refleja características de sustentantes | | Refleja contexto cultural de sustentantes | | Provoca sesgo | |
| | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % |
| R1 | | | | | | | | | 5/6 | 83.3 | | | | | 5/6 | 83.3 | 0/6 | 0 |
| R2 | | | | | | | | | | | | | | | | | 0/6 | 0 |
| R3 | | | | | | | | | | | | | | | | | 0/6 | 0 |
| R4 | | | 5/6 | 83.3 | 5/6 | 83.3 | | | | | | | | | | | 0/6 | 0 |
| R5 | | | | | | | | | | | | | | | | | 1/6 | 16.7 |
| R6 | | | | | | | | | | | | | | | | | 0/6 | 0 |
| R7 | | | | | | | | | | | | | | | | | 0/6 | 0 |
| R8 | | | | | | | | | | | | | | | | | 0/6 | 0 |
| R9 | | | | | | | | | | | | | | | | | 0/6 | 0 |
| R10 | | | | | | | | | | | | | | | | | 0/6 | 0 |
| R11 | | | | | | | | | | | | | | | | | 0/6 | 0 |
| R12 | | | 5/6 | 83.3 | | | | | | | | | | | | | 0/6 | 0 |
| R13 | | | | | | | | | | | | | | | | | 0/6 | 0 |
| R14 | | | | | | | | | | | 4/6 | 66.7 | | | | | 0/6 | 0 |
| R15 | | | | | | | | | | | | | | | | | 0/6 | 0 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R16 | | | | | | | | | | | | | | | | | 0/6 | 0 |
| R17 | | | | | | | 4/6 | 66.7 | 4/6 | 66.7 | 5/6 | 83.3 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 |
| R18 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | 6/6 | | 5/6 | 83.3 | | | 4/6 | 66.7 | 0/6 | 0 |
| R19 | | | | | | | | | | | | | | | 5/6 | 83.3 | 0/6 | 0 |
| R20 | | | | | | | | | | | | | | | | | 0/6 | 0 |
| R21 | | | | | | | | | | | | | | | | | 2/6 | 33.3 |
| R22 | | | | | | | | | | | | | | | | | 0/6 | 0 |
| R23 | | | | | | | | | | | | | | | | | 0/6 | 0 |
| R24 | | | | | | | | | | | 5/6 | 83.3 | | | 4/6 | 66.7 | 4/6 | 66.7 |
| R25 | 5/6 | 83.3 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | | | | | 4/6 | 66.7 | 4/6 | 66.7 | 0/6 | 0 |
| R26 | 3/6 | 50 | 5/6 | 83.3 | 3/6 | 50 | 2/6 | 33.3 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | 0/6 | 0 |
| R27 | | | | | | | 5/6 | 83.3 | | | | | | | 4/6 | 66.7 | 2/6 | 33.3 |
| R28 | | | | | | | 5/6 | 83.3 | | | | | | | 4/6 | 66.7 | 4/6 | 66.7 |
| R29 | | | | | | | 5/6 | 83.3 | | | | | | | | | 0/6 | 0 |
| R30 | | | | | | | 5/6 | 83.3 | | | | | | | | | 0/6 | 0 |
| R31 | | | | | | | 5/6 | 83.3 | | | | | | | 4/6 | 66.7 | 2/6 | 33.3 |
| R32 | | | | | | | 5/6 | 83.3 | | | | | | | | | 0/6 | 0 |

FR= frecuencia relativa con N=6 jueces.

Las celdas en blanco representan consenso de los jueces respecto a la respuesta "De acuerdo" con la aseveración que define cada aspecto evaluado.

Cuando los jueces no llegaron a consensos se expresan las frecuencias relativas correspondientes (5/6, 4/6, 3/6, 2/6 ó 1/6) y su valor en porcentajes.

Tabla A5.2. Frecuencias de los acuerdos entre los panelistas en los aspectos evaluados de los Nodos y Características de los reactivos de Gramática.

| Reactivos | NODOS | | | | | | | | CARACTERISTICAS DE LOS REACTIVOS | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Relevante | | Representativo | | Pertinente | | Claramente descrito | | Similar a la realidad | | Nivel apropiado | | Refleja características de sustentantes | | Refleja contexto cultural de sustentantes | | Provoca sesgo | |
| | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % |
| R33 | | | | | | | | | | | | | | | | | 0/4 | 0 |
| R34 | | | | | | | | | | | | | | | | | 1/4 | 25 |
| R35 | | | | | | | | | | | | | | | | | 0/4 | 0 |
| R36 | | | | | | | | | | | | | | | 2/4 | 50 | 0/4 | 0 |
| R37 | | | | | | | | | | | | | | | | | 1/4 | 25 |
| R38 | | | | | | | | | | | | | | | | | 0/4 | 0 |
| R39 | | | | | | | | | | | | | | | | | 0/4 | 0 |
| R40 | | | | | | | | | | | | | | | | | 0/4 | 0 |
| R41 | | | | | | | | | | | | | | | 3/4 | 75 | 0/4 | 0 |
| R42 | | | | | | | | | | | | | | | | | 0/4 | 0 |
| R43 | | | | | | | | | | | | | | | | | 0/4 | 0 |
| R44 | | | | | | | | | | | | | | | | | 0/4 | 0 |
| R45 | | | | | | | | | | | | | | | | | 1/4 | 25 |
| R46 | | | | | | | | | | | | | | | | | 1/4 | 25 |
| R47 | | | | | | | | | | | | | | | 3/4 | 75 | 1/4 | 25 |
| R48 | | | | | | | | | | | | | | | 1/4 | 25 | 1/4 | 25 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R49 | | | | | | | | | | | | | | | | 0/4 | 0 |
| R50 | | | | | | | | | | | | | | | | 0/4 | 0 |
| R51 | | | | | | | | | | | | | | 3/4 | 75 | 0/4 | 0 |
| R52 | | | | | | | | | | | | | | 3/4 | 75 | 0/4 | 0 |
| R53 | | | | | | | | | | | | | | 3/4 | 75 | 0/4 | 0 |
| R54 | | | | | | | | | | | | | | 3/4 | 75 | 1/4 | 25 |
| R55 | | | | | | | | | | | | | | 3/4 | 75 | 1/4 | 25 |
| R56 | | | | | | | | | | | | | | | | 1/4 | 25 |
| R57 | | | | | | | | | | | | | | | | 0/4 | 0 |
| R58 | | | | | | | | | | | | | | 3/4 | 75 | 0/4 | 0 |
| R59 | | | | | | | | | | | | | | | | 0/4 | 0 |
| R60 | | | | | | | | | | | | | | | | 1/4 | 25 |
| R61 | | | | | | | | | | | | | | | | 1/4 | 25 |
| R62 | | | | | | | | | | | | | | | | 1/4 | 25 |
| R63 | | | | | | | | | | | | | | | | 1/4 | 25 |
| R64 | | | | | | | | | | | | | | | | 1/4 | 25 |
| R65 | | | | | | | | | | | | | | | | 0/4 | 0 |
| R66 | | | | | | | | | | | | | | | | 0/4 | 0 |
| Total | - | - | - | - | - | - | - | - | - | - | - | - | - | 10/34 | 29.4 | 29.4 | |

FR= frecuencia relativa con N=4 jueces.

Cuando los jueces no llegaron a consensos se expresan las frecuencias relativas correspondientes (3/4, 2/4 ó 1/4) y su valor en porcentajes, indicando la posibilidad de que el reactivo no pertenezca al DULM, o que podría tener problemas de autenticidad, interactividad o sesgo. Las celdas en blanco representan 100% de acuerdo entre los jueces.

Tabla A53. Frecuencias de los acuerdos entre los panelistas en los aspectos evaluados de los Nodos y Características de los reactivos de Lectura.

| Reactivos | LECTURA | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | NODOS | | | | | | | | CARACTERISTICAS DE LOS REACTIVOS | | | | | | | | | |
| | Relevante | | Representativo | | Pertinente | | Claramente descrito | | Similar a la realidad | | Nivel apropiado | | Refleja características de sustentantes | | Refleja contexto cultural de sustentantes | | Provoca sesgo | |
| | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % |
| R67 | | | 5/6 | 83.3 | | | | | | | | | | | | | 1/6 | 16.7 |
| R68 | | | | | | | | | | | | | | | | | 1/6 | 16.7 |
| R69 | | | | | | | | | | | | | | | | | 1/6 | 16.7 |
| R70 | | | | | | | | | | | | | | | | | 1/6 | 16.7 |
| R71 | | | | | | | | | | | | | | | | | 1/6 | 16.7 |
| R72 | | | | | | | 5/6 | 83.3 | | | 5/6 | 83.3 | | | | | 1/6 | 16.7 |
| R73 | | | | | | | | | | | | | | | | | 1/6 | 16.7 |
| R74 | | | | | | | | | | | | | | | | | 1/6 | 16.7 |
| R75 | | | | | | | | | | | 5/6 | 83.3 | | | | | 2/6 | 33.3 |
| R76 | | | | | | | | | | | | | | | | | 1/6 | 16.7 |
| R77 | | | | | | | | | 5/6 | 83.3 | | | | | | | 3/6 | 50 |
| R78 | 5/6 | 83.3 | 4/6 | 66.7 | | | 5/6 | 83.3 | 5/6 | 83.3 | 4/6 | 66.7 | 5/6 | 83.3 | 4/6 | 66.7 | 2/6 | 33.3 |
| R79 | | | | | | | | | 5/6 | 83.3 | 5/6 | 83.3 | | | | | 1/6 | 16.7 |
| R80 | 5/6 | 83.3 | 5/6 | 83.3 | | | | | 5/6 | 83.3 | 5/6 | 83.3 | | | | | 1/6 | 16.7 |
| R81 | | | | | | | | | | | | | | | | | 1/6 | 16.7 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R82 | | | | | | | | | | | 4/6 | 66.7 | | | 4/6 | 66.7 | 3/6 | 50 |
| R83 | | | | | | | | | | | | | | | | | 3/6 | 50 |
| R84 | | | | | | | | | 5/6 | 83.3 | | | | | | | 1/6 | 16.7 |
| R85 | 5/6 | 83.3 | 5/6 | 83.3 | | | 4/6 | 66.7 | 5/6 | 83.3 | | | | | 3/6 | 50 | 3/6 | 50 |
| R86 | | | 5/6 | 83.3 | | | 5/6 | 83.3 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | 3/6 | 50 | 3/6 | 50 |
| R87 | | | | | | | | | | | | | | | 5/6 | 83.3 | 1/6 | 16.7 |
| R88 | | | | | | | | | | | | | | | 5/6 | 83.3 | 1/6 | 16.7 |
| R89 | | | 4/6 | 66.7 | 5/6 | 83.3 | 5/6 | 83.3 | 5/6 | 83.3 | 5/6 | 83.3 | 5/6 | 83.3 | 4/6 | 66.7 | 3/6 | 50 |
| R90 | 5/6 | 83.3 | 5/6 | 83.3 | | | | | | | 4/6 | 66.7 | 5/6 | 83.3 | 4/6 | 66.7 | 4/6 | 66.7 |
| R91 | | | | | | | | | | | | | | | | | 3/6 | 50 |
| R92 | | | | | | | | | | | | | | | | | 1/6 | 16.7 |
| R93 | | | | | | | | | | | | | | | | | 1/6 | 16.7 |
| R94 | | | | | | | | | | | | | | | 4/6 | 66.7 | 3/6 | 50 |
| R95 | | | | | | | | | | | | | | | 4/6 | 66.7 | 1/6 | 16.7 |
| R96 | | | | | | | 5/6 | 83.3 | 6/6 | 100 | 5/6 | 83.3 | | | 4/6 | 66.7 | 1/6 | 16.7 |
| R97 | | | | | | | | | | | | | | | | | 1/6 | 16.7 |
| R98 | | | | | | | | | | | | | | | | | 1/6 | 16.7 |
| R99 | | | | | | | | | | | | | | | | | 1/6 | 16.7 |
| R100 | | | | | | | | | | | | | | | | | 1/6 | 16.7 |

FR: frecuencia relativa con N=6 jueces. Cuando los jueces no llegaron a consensos se expresan las frecuencias relativas correspondientes (5/6, 4/6, 3/6, 2/6 ó 1/6) y su valor en porcentajes, indicando la posibilidad de que el reactivo no pertenezca al DULM, o que podría tener problemas de autenticidad, interactividad o sesgo. Las celdas en blanco representan 100% de acuerdo entre los jueces.

Tabla A6.4. Frecuencias de los acuerdos entre los panelistas en los aspectos evaluados de las habilidades/competencias de los reactivos de Comprensión auditiva.

| | Registro | | Opinión/ hecho | | Pregunta | | Nativos | | Inconve- nientes | | Instruc/ tareas | | Puntos de vista | | Info personal | | Solicita ayuda | | Inter/ Opinión | | Abstrac/ cultural | | Argumento | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % |
| 1 | | | | | 6/6 | [] | | | | | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | | | | | | |
| 2 | | | | | | | | | | | | | | | | | | | | | | | 6/6 | [] |
| 3 | | | | | | | | | | | | | | | | | 6/6 | [] | | | | | 6/6 | [] |
| 4 | | | | | | | | | | | | | | | | | | | | | | | 6/6 | [] |
| 5 | | | | | | | | | | | | | 6/6 | [] | | | | | | | | | | |
| 6 | | | | | | | | | | | | | | | | | | | | | | | 6/6 | [] |
| 7 | | | | | | | | | | | | | | | | | | | | | | | 6/6 | [] |
| 8 | | | | | | | | | 6/6 | [] | | | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | | | | | | | [] | | | | | | |
| 10 | | | | | | | | | | | | | | | | | | | | | | | 6/6 | [] |
| 11 | | | | | | | | | | | | | 6/6 | [] | | | | | 6/6 | [] | | | | |
| 12 | | | | | | | | | | | | | | | | | | | | | | | 4/6 | 66.7 |
| 13 | | | 6/6 | [] | | | | | | | 6/6 | [] | | | | | | | | | | | | |
| 14 | | | | | | | | | 2/6 | 33.3 | 2/6 | 33.3 | | | | | | | | | | | 2/6 | 33.3 |
| 15 | | | | | | | | | | | | | | | | | | | | | | | 6/6 | [] |
| 16 | | | | | | | | | 6/6 | [] | | | | | | | | | | | | | | |
| 17 | | | | | | | | | | | 2/6 | 33.3 | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 18 | 4/6 | 66.7 | 4/6 | 66.7 | | | | | 4/6 | 66.7 | | | | | | | | | 4/6 | 66.7 | | | 4/6 | 66.7 |
| 19 | | | | | | | | | | | | | | | | | | | | | | | | |
| 20 | | | | | | | | | | | | | | | | | | | | | | | 4/6 | 66.7 |
| 21 | | | | | | | | | | | 6/6 | [] | | | | | | | | | | | | |
| 22 | | | | | | | | | | | | | | | | | | | | | 6/6 | [] | | |
| 23 | | | | | | | | | | | | | 6/6 | [] | | | | | | | | | | |
| 24 | | | | | | | | | | | 2/6 | 33.3 | | | | | | | 6/6 | [] | | | | |
| 25 | | | | | | | | | | | | | | | | | | | | | 2/6 | 33.3 | 2/6 | 33.3 |
| 26 | | | 4/6 | 66.7 | | | | | 4/6 | 66.7 | | | 4/6 | 66.7 | | | | | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 |
| 27 | | | | | | | | | 6/6 | [] | | | 6/6 | [] | | | | | | | | | | |
| 28 | | | | | | | | | 2/6 | 33.3 | | | | | | | | | | | 6/6 | [] | | |
| 29 | | | | | | | | | | | | | | | | | | | | | | | 6/6 | [] |
| 30 | | | 6/6 | [] | | | | | | | | | | | | | | | | | | | | |
| 31 | | | | | | | | | | | | | | | | | | | | | 6/6 | [] | [] | |
| 32 | | | | | | | | | 6/6 | [] | | | | | | | | | | | | | | |

FR= frecuencia relativa con N=6 jueces.

[] = consenso de que el reactivo mide la competencia para la que fue diseñado el reactivo.

Cuando los jueces no llegaron a consensos se expresan las frecuencias relativas correspondientes (FR: 4/6, 3/6 ó 2/6) y su valor en porcentajes. Acuerdos menores al 75% indican que el reactivo no mide la habilidad/competencia para la que fue diseñado.

Las celdas en blanco representan consensos para respuestas marcadas con 0 (en desacuerdo).

Tabla A5.5. Frecuencias de los acuerdos entre los panelistas en los aspectos evaluados de las habilidades/competencias de los reactivos de Gramática.

| Reactivos | Sustan-Tivos | | Pose-sivos | | Arts demos-trativos | | Adj Ble ous | | Adj Compa-rativos | | Adver-bios | | Superla-Tivos | | Pronom person | | Preposi-ciones | | Conjun-ciones | | Nega-tivos | | Who What | | Verbos | | Tiempos gramat | | Verbos aux | | Verbos modales | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % |
| 33 | 2/4 | 50 | | | | | | | | | | | | | | | | | | | 4 | ¡ | 4 | ¡ | 1/4 | 25 | 4 | ¡ | | | | |
| 34 | ¾ | 75 | | | 1/4 | 25 | | | | | ¼ | 25 | | | 3/4 | 75 | 1/4 | 25 | | | | | | | 1/4 | 25 | 4 | ¡ | 1/4 | 25 | | |
| 35 | 2/4 | 50 | | | 2/4 | 50 | | | | | | | | | 1/4 | 25 | | | | | | | 1/4 | *25 | 4 | ¡ | 4/4 | [] | 4/4 | ¡ | | |
| 36 | 2/4 | 50 | | | | | | | | | | | | | 2/4 | 50 | | | | | | | | | 3/4 | 75 | 4/4 | [] | 4/4 | [] | | |
| 37 | 2/4 | 50 | | | | | | | | | | | | | | | 1/4 | 25 | | | | | | | | | 4/4 | [] | 4/4 | [] | | |
| 38 | 2/4 | 50 | | | | | | | | | | | | | 2/4 | 50 | | | | | | | | | 1/4 | 25 | 4/4 | [] | 4/4 | [] | | |
| 39 | 2/4 | 50 | | | 1/4 | 25 | | | | | | | | | 2/4 | 50 | | | | | | | | | 1/4 | 25 | 4/4 | [] | 4/4 | [] | | |
| 40 | ¾ | 75 | | | 2/4 | 50 | | | 1/4 | 25 | | | | | 3/4 | 75 | | | | | | | | | | | 4/4 | [] | 4/4 | [] | | |
| 41 | 2/4 | 50 | | | 1/4 | 25 | | | | | | | | | | | | | | | | | | | | | 4/4 | [] | 3/4 | 75 | | |
| 42 | 2/4 | 50 | | | | | | | | | 4/4 | [] | | | | | | | | | | | | | | | | | | | | |
| 43 | 2/4 | 50 | | | 1/4 | 25 | | | | | 4/4 | [] | | | | | | | | | | | | | | | | | | | | |
| 44 | 2/4 | 50 | | | | | | | | | 4/4 | [] | | | | | | | | | | | | | | | | | | | | |
| 45 | 2/4 | 50 | | | | | | | | | 4/4 | [] | | | 1/4 | 25 | | | | | | | | | 1/4 | 25 | 1/4 | 25 | 2/4 | 50 | | |
| 46 | 2/4 | 50 | | | 1/4 | 25 | | | | | 4/4 | [] | | | | | | | | | | | | | 1/4 | 25 | 2/4 | 50 | | | | |
| 47 | 2/4 | 50 | | | 1/4 | 25 | | | | | | | | | | | | | | | | | | | 3/4 | 75 | 4/4 | [] | | | | |
| 48 | 2/4 | 50 | | | 1/4 | 25 | | | | | | | | | | | | | | | | | | | 3/4 | 75 | 4/4 | [] | | | | |
| 49 | 2/4 | 50 | | | 1/4 | 25 | | | | | | | | | | | | | | | | | | | 3/4 | 75 | 4/4 | [] | | | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 50 | 2/4 | 50 | | | 1/4 | 25 | | | | | | | | | | | | | | | | | | | 4 | [] | 4/4 | [] | | | | |
| 51 | 2/4 | 50 | | | 1/4 | 25 | | | | | | | | | | | | | | | | | | | 4/4 | [] | 4/4 | [] | | | | |
| 52 | 2/4 | 50 | | | 1/4 | 25 | | | | | | | | | | | | | | | | | | | 4/4 | [] | 4/4 | [] | 4/4 | [] | | |
| 53 | 2/4 | 50 | | | 0 | 0 | | | | | | | | | | | | | | | | | | | 4/4 | [] | 4/4 | [] | 4/4 | [] | | |
| 54 | 2/4 | 50 | | | 0 | 0 | | | | | | | | | | | | | | | | | | | 4/4 | [] | 4/4 | [] | 4/4 | [] | | |
| 55 | 2/4 | 50 | | | 1/4 | 25 | | | | | | | | | | | | | | | | | | | 4/4 | [] | 4/4 | [] | 4/4 | [] | | |
| 56 | 2/4 | 50 | | | 1/4 | 25 | | | 1/4 | 25 | | | | | | | | | | | | | | | 4/4 | [] | 4/4 | [] | 4/4 | [] | | |
| 57 | 2/4 | 50 | | | 1/4 | 25 | | | 4/4 | [] | | | 3/4 | *75 | | | | | | | | | | | 1/4 | 25 | | | | | | |
| 58 | 2/4 | 50 | | | 1/4 | 25 | ¼ | 25 | 4/4 | [] | | | 4/4 | [] | | | | | | | | | | | 1/4 | 25 | | | | | | |
| 59 | 2/4 | 50 | | | 1/4 | 25 | ¼ | 25 | 4 | ¡ | | | 4/4 | [] | | | | | | | | | | | 1/4 | 25 | | | | | | |
| 60 | 2/4 | 50 | | | | | 2/4 | 50 | 3/4 | 75 | | | 4/4 | [] | | | | | | | | | | | 1/4 | 25 | | | | | | |
| 61 | 2/4 | 50 | | | | | | | | | | | | | 1/4 | 25 | | | | | | | | | | | | | | | | |
| 62 | 2/4 | 50 | | | 1/4 | 25 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 63 | 2/4 | 50 | | | 1/4 | 25 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 64 | 2/4 | 50 | | | | | | | | | | | | | | | | | 1/4 | 25 | | | | | | | | | | | | |
| 65 | 2/4 | 50 | | | | | | | | | | | | | | | | | | | | | | | | | 4/4 | [] | | | | |
| 66 | 2/4 | 50 | | | | | | | | | | | | | 1/4 | 25 | | | | | | | | | | | 4/4 | [] | | | | |

FR= frecuencia relativa con N = 4 jueces.

[] = consenso de que el reactivo mide la competencia para la que fue diseñado.

Cuando los jueces no llegaron a consensos se expresan las frecuencias relativas correspondientes (FR: 3/4, 2/4 ó 1/4) y su valor en porcentajes. Acuerdos menores al 75% indican que el reactivo no mide la habilidad/competencia para la que fue diseñado.

4 ¡ = consenso de que el reactivo mide una competencia para la cual no fue diseñado.

* = acuerdo en que el reactivo mide una competencia para la cual no fue diseñado.

Las celdas en blanco representan consenso para respuestas marcadas con 0 (en desacuerdo).

Tabla A5.6. Frecuencias de los acuerdos entre los panelistas en los aspectos evaluados de las habilidades/competencias de los reactivos de Lectura.

| | Info rutinaria | | Info no rutinaria | | Cartas y notas temas conocidos | | Instrucciones catálogos | | Notas solicitudes | | Periódicos hoteles | | Opinión personal | | Ensayo temas conocidas | | Apuntes apoyo lengua materna | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % | FR | % |
| 67 | 6/6 | [] | 6/6 | [] | 4/6 | 66.7 | 2/6 | 33.3 | 6/6 | [] | 6/6 | [] | 4/6 | 66.7 | 6/6 | [] | 6/6 | [] |
| 68 | 6/6 | [] | 6/6 | [] | 4/6 | 66.7 | 1/6 | 16.7 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 4/6 | 66.7 | 4/6 | 66.7 |
| 69 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 1/6 | 16.7 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 4/6 | 66.7 | 4/6 | 66.7 |
| 70 | 6/6 | [] | 6/6 | [] | 4/6 | 66.7 | 1/6 | 16.7 | 4/6 | 66.7 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |
| 71 | 6/6 | [] | 6/6 | [] | 4/6 | 66.7 | 1/6 | 16.7 | 4/6 | 66.7 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |
| 72 | 6/6 | [] | 6/6 | [] | 4/6 | 66.7 | 1/6 | 16.7 | 4/6 | 66.7 | 6/6 | [] | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 |
| 73 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 1/6 | 16.7 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 |
| 74 | 6/6 | [] | 6/6 | [] | 4/6 | 66.7 | 1/6 | 16.7 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 |
| 75 | 6/6 | [] | 6/6 | [] | 4/6 | 66.7 | 1/6 | 16.7 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 |
| 76 | 6/6 | [] | 6/6 | [] | 4/6 | 66.7 | 1/6 | 16.7 | 4/6 | 66.7 | 6/6 | [] | 4/6 | 66.7 | 4/6 | 66.7 | 4/6 | 66.7 |
| 77 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 1/6 | 16.7 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |
| 78 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 1/6 | 16.7 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |
| 79 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 5/6 | 83.3 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |
| 80 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 1/6 | 16.7 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |
| 81 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 1/6 | 16.7 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |
| 82 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 1/6 | 16.7 | 4/6 | 66.7 | 6/6 | [] | 4/6 | 66.7 | 6/6 | [] | 6/6 | [] |
| 83 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 1/6 | 16.7 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 84 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 3/6 | 50 | 6/6 | [] | 6/6 | [] | 3/6 | 50 | 6/6 | [] | 6/6 | [] |
| 85 | 6/6 | [] | 4/6 | 66.7 | 4/6 | 66.7 | 1/6 | 16.7 | 4/6 | 66.7 | 6/6 | [] | 6/6 | [] | 4/6 | 66.7 | 6/6 | [] |
| 86 | 6/6 | [] | 4/6 | 66.7 | 4/6 | 66.7 | 1/6 | 16.7 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |
| 87 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 1/6 | 16.7 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |
| 88 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 1/6 | 16.7 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |
| 89 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 1/6 | 16.7 | 1/6 | 16.7 | 1/6 | 16.7 | 6/6 | [] | 6/6 | [] | 6/6 | [] |
| 90 | 6/6 | [] | 6/6 | [] | 6/6 | 66.7 | 1/6 | 16.7 | 6/6 | [] | 4/6 | 66.7 | 6/6 | [] | 6/6 | [] | 4/6 | 66.7 |
| 91 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 3/6 | 50 | 6/6 | [] | 6/6 | [] | 3/6 | 50 | 6/6 | [] | 6/6 | [] |
| 92 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 5/6 | 83.3 | 6/6 | [] | 6/6 | [] | 3/6 | 50 | 6/6 | [] | 6/6 | [] |
| 93 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 3/6 | 50 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |
| 94 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 4/6 | 66.7 | 6/6 | [] | 6/6 | [] | 3/6 | 50 | 6/6 | [] | 4/6 | 66.7 |
| 95 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 2/6 | 33.3 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |
| 96 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 3/6 | 50 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |
| 97 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 3/6 | 50 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |
| 98 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 3/6 | 50 | 6/6 | [] | 6/6 | [] |
| 99 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 5/6 | 83.3 | 6/6 | [] | 6/6 | [] | 3/6 | 50 | 6/6 | [] | 6/6 | [] |
| 100 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 5/6 | 83.3 | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] | 6/6 | [] |

Fr: frecuencia relativa con N=6 jueces.[] = consenso de que el reactivo mide la competencia para la que fue diseñado.

Cuando los jueces no llegaron a consensos se expresan las frecuencias relativas correspondientes (5/6, 4/6, 3/6 ó 2/6) y su valor en porcentajes. Acuerdos menores al 75% indican que el reactivo no mide la habilidad/competencia para la que fue diseñado.

No hubo consenso para respuestas marcadas con 0 (en desacuerdo).

## ANEXO 6.1. VALIDACION DE CONSTRUCTO
## Análisis de la información de salida del patrón de respuestas al EXEDII. Modelamiento Rasch

**Población analizada.** Los datos analizados corresponden a los resultados obtenidos en EXEDII por los alumnos de las cohortes de 2005 y 2006, las cuales fueron analizadas con modelamiento Rasch, mediante el programa Winsteps.
Se presentan a continuación los pasos seguidos en la interpretación:

Para analizar datos en el modelo Rasch es necesario preparar de antemano dos archivos: archivo de datos, que consiste en la base de datos que van a ser analizados, los cuales deberán estar capturados en un solo archivo, con extensión txt. En este caso, los datos son las respuestas a los cien reactivos del EXEDII en igual número de columnas y 2260 sustentantes en los renglones. El archivo de control se adaptó del utilizado por González Montesinos (2004) en su tesis doctoral y consiste en una serie de instrucciones en las que se especifican las características del análisis que se desea realizar. El archivo de control se muestra a continuación:

**Paso 1.** Correr el análisis con Winsteps con el siguiente Archivo de Control:

```
;WINSTEPS CONTROL FILE FOR EXEDII
&INST
TITLE="Analisis Defintivo EXEDII"
NI=100
ITEM1=9
NAME1=1
NAMELEN=8
IDELQ= Y
CODES = 0 1
KEY=11111111111111111111111111111111111111111111111111111111111111111111111111111111
11111111111111111111111111
XWIDE=1
LOCAL=Y
PTBIS=Y
DISC=Y
DATA= H:\Virginia Velasco\Doctorado\RASCH\RASCH Definitivo\DEFIN1.dat
&END
R1 ;Item1
R2 ;Item2
R3 ;Item3
R4 ;Item4
R5 ;Item5
R6 ;Item6
R7 ;Item7
R8 ;Item8
R9 ;Item9
R10 ;Item10
R11 ;Item11
R12 ;Item12
```

R13 ;Item13
R14 ;Item14
R15 ;Item15
R16 ;Item16
R17 ;Item17
R18 ;Item18
R19 ;Item19
R20 ;Item20
R21 ;Item21
R22 ;Item22
R23 ;Item23
R24 ;Item24
R25 ;Item25
R26 ;Item26
R27 ;Item27
R28 ;Item28
R29 ;Item29
R30 ;Item30
R31 ;Item31
R32 ;Item32
R33 ;Item33
R34 ;Item34
R35 ;Item35
R36 ;Item36
R37 ;Item37
R38 ;Item38
R39 ;Item39
R40 ;Item40
R41 ;Item41
R42 ;Item42
R43 ;Item43
R44 ;Item44
R45 ;Item45
R46 ;Item46
R47 ;Item47
R48 ;Item48
R49 ;Item49
R50 ;Item50
R51 ;Item51
R52 ;Item52
R53 ;Item53
R54 ;Item54
R55 ;Item55
R56 ;Item56
R57 ;Item57
R58 ;Item58
R59 ;Item59

R60 ;Item60
R61 ;Item61
R62 ;Item62
R63 ;Item63
R64 ;Item64
R65 ;Item65
R66 ;Item66
R67 ;Item67
R68 ;Item68
R69 ;Item69
R70 ;Item70
R71 ;Item71
R72 ;Item72
R73 ;Item73
R74 ;Item74
R75 ;Item75
R76 ;Item76
R77 ;Item77
R78 ;Item78
R79 ;Item79
R80 ;Item80
R81 ;Item81
R82 ;Item82
R83 ;Item83
R84 ;Item84
R85 ;Item85
R86 ;Item86
R87 ;Item87
R88 ;Item88
R89 ;Item89
R90 ;Item90
R91 ;Item91
R92 ;Item92
R93 ;Item93
R94 ;Item94
R95 ;Item95
R96 ;Item96
R97 ;Item97
R98 ;Item98
R99 ;Item99
R100 ;Item100

**Paso 2.** Al correr el programa se obtuvo la Tabla de convergencia que aparece a continuación:

```
     2260 PERSON       Records Input.
                          CONVERGENCE TABLE
+Control: H:\Virginia--------------------Output: H:\Virginia-----------------+
|    PROX          ACTIVE COUNT       EXTREME 5 RANGE      MAX LOGIT CHANGE  |
| ITERATION   PERSONS ITEMS   CATS     PERSONS ITEMS      MEASURES  STRUCTURE|
>=============================================================================<
|        1     2260     100     2       6.23    2.87        4.5951           |
>=============================================================================<
|        2     2260     100     2       6.83    3.40         .4425           |
>=============================================================================<
|        3     2260     100     2       7.05    3.50         .1666           |
|Control: H:\Virginia--------------------Output: H:\Virginia-----------------|
|    UCON      MAX SCORE   MAX LOGIT    LEAST CONVERGED     CATEGORY STRUCTURE|
| ITERATION   RESIDUAL*    CHANGE    PERSON ITEM     CAT   RESIDUAL   CHANGE |
>=============================================================================<
|       1       -21.55      -.1872      40     86*                           |
>=============================================================================<
|       2        -8.72      -.0478     219     86*                           |
>=============================================================================<
|       3        -3.49      -.0112     225     86*                           |
>=============================================================================<
|       4        -1.50       .0032     225     86*                           |
+-----------------------------------------------------------------------------+
  Calculating Fit Statistics
>=============================================================================<
Standardized Residuals N(0,1)  Mean: .01 S.D.: 1.00
Analisis Defintivo EXEDII
+-----------------------------------------------------------------------------+
| PERSONS    2260 INPUT    2260 MEASURED              INFIT         OUTFIT    |
|           SCORE     COUNT     MEASURE   ERROR     IMNSQ   ZEMP  OMNSQ   ZEMP|
| MEAN       57.0     100.0         .45     .25      1.00    -.1   1.00     .0|
| S.D.       20.3       .0         1.17     .06       .09    1.0    .21    1.0|
| REAL RMSE    .26  ADJ.SD    1.14  SEPARATION  4.37  PERSON RELIABILITY  .95|
|-----------------------------------------------------------------------------|
| ITEMS     100 INPUT     100 MEASURED               INFIT         OUTFIT    |
| MEAN    1289.3    2260.0         .00     .05      1.00    -.1   1.00     .0|
| S.D.     361.0       .0         .90     .00       .13    1.0    .26    1.0|
| REAL RMSE    .05  ADJ.SD     .90  SEPARATION 17.25  ITEM   RELIABILITY 1.00|
+-----------------------------------------------------------------------------+
 Output written to H:\Virginia Velasco\Doctorado\RASCH\RASCH Definitivo\ZOU225ws.txt
 CODES ="0 1"
Measures constructed: use "Diagnosis" and "Output Tables" menus
  Processing Table 13
  Building Category/Option/Distractor Table 13
>=============================================================================<
```

**Paso 3.** Desde el menú principal se seleccionaron las tablas de salida para interpretar el análisis. Los criterios para la interpretación fueron consultados en el artículo de González Montesinos (2007) y directamente del manual de Winsteps (1991-2006).

Primeramente se presentan las estadísticas descriptivas **sumarias** de los **sustentantes**.

```
TABLE 3.1 Analisis Defintivo EXEDII              ZOU225ws.txt Sep  3 13:17 2007
INPUT: 2260 PERSONS, 100 ITEMS  MEASURED: 2260 PERSONS, 100 ITEMS, 2 CATS   3.37
-------------------------------------------------------------------------------

     SUMMARY OF 2260 MEASURED PERSONS
+-----------------------------------------------------------------------------+
|          RAW                          MODEL         INFIT        OUTFIT     |
|         SCORE     COUNT     MEASURE   ERROR      MNSQ   ZEMP   MNSQ   ZEMP |
|-----------------------------------------------------------------------------|
| MEAN     57.0     100.0         .45     .25      1.00    -.1   1.00     .0 |
| S.D.     20.3       .0        1.17     .06       .09    1.0    .21    1.0 |
| MAX.     99.0     100.0        4.96    1.01      1.35    3.7   3.23    3.7 |
| MIN.     14.0     100.0       -2.09     .22       .73   -3.9    .36   -3.4 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .26  ADJ.SD    1.14  SEPARATION  4.37  PERSON RELIABILITY  .95 |
|MODEL RMSE    .26  ADJ.SD    1.14  SEPARATION  4.44  PERSON RELIABILITY  .95 |
| S.E. OF PERSON MEAN = .02                                                   |
+-----------------------------------------------------------------------------+
PERSON RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) PERSON RAW SCORE RELIABILITY = .95

     SUMMARY OF 100 MEASURED ITEMS
+-----------------------------------------------------------------------------+
|          RAW                          MODEL         INFIT        OUTFIT     |
|         SCORE     COUNT     MEASURE   ERROR      MNSQ   ZEMP   MNSQ   ZEMP |
|-----------------------------------------------------------------------------|
| MEAN   1289.3    2260.0         .00     .05      1.00    -.1   1.00     .0 |
| S.D.    361.0       .0         .90     .00       .13    1.0    .26    1.0 |
| MAX.   1999.0    2260.0        1.80     .07      1.34    2.6   2.42    2.5 |
| MIN.    571.0    2260.0       -2.01     .05       .77   -2.2    .58   -2.0 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .05  ADJ.SD     .90  SEPARATION 17.25  ITEM   RELIABILITY 1.00 |
|MODEL RMSE    .05  ADJ.SD     .90  SEPARATION 17.67  ITEM   RELIABILITY 1.00 |
| S.E. OF ITEM MEAN = .09                                                     |
+-----------------------------------------------------------------------------+
UMEAN=.000 USCALE=1.000
ITEM RAW SCORE-TO-MEASURE CORRELATION = -1.00
```

**Paso 4.** Revisión de la tabla que muestra la calibración de los reactivos y los estadígrafos de ajuste al modelo.

```
TABLE 13.1 Analisis Defintivo EXEDII              ZOU225ws.txt Sep  3 13:17
2007
INPUT: 2260 PERSONS, 100 ITEMS  MEASURED: 2260 PERSONS, 100 ITEMS, 2 CATS
3.37
-------------------------------------------------------------------------------
-
PERSON: REAL SEP.: 4.37  REL.: .95 ... ITEM: REAL SEP.: 17.25  REL.: 1.00

          ITEMS STATISTICS:  MEASURE ORDER

+-----------------------------------------------------------------------------+
|ENTRY    RAW                          |   INFIT  |  OUTFIT  |PTBIS|      |      |
|NUMBER  SCORE  COUNT  MEASURE  ERROR|MNSQ  ZEMP|MNSQ  ZEMP|CORR.|DISCR| ITEMS|
|------------------------------------+----------+----------+-----+-----+------|
|    86    571   2260     1.80    .05|1.23   1.2|1.43   1.5|  .24|  .67| R86  |
|    79    611   2260     1.68    .05|1.20   1.1|1.39   1.5|  .27|  .69| R79  |
|    91    622   2260     1.65    .05|1.02    .1|1.16    .7|  .40|  .93| R91  |
|    20    625   2260     1.64    .05|1.03    .2|1.19    .8|  .37|  .91| R20  |
|   100    672   2260     1.51    .05| .90   -.7| .88   -.6|  .52| 1.15| R100 |
|    38    729   2260     1.35    .05|1.10    .6|1.21   1.0|  .34|  .81| R38  |
|    96    733   2260     1.34    .05|1.02    .2|1.03    .2|  .43|  .96| R96  |
|    83    746   2260     1.31    .05| .88   -.8| .88   -.7|  .54| 1.18| R83  |
|    66    749   2260     1.30    .05|1.00    .0|1.06    .3|  .43|  .97| R66  |
|    25    780   2260     1.22    .05| .90   -.7| .94   -.3|  .53| 1.15| R25  |
```

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | 782 | 2260 | 1.21 | .05 | 1.28 | 1.8 | 1.40 | 1.9 | .23 | .49 | R12 |
| 5 | 787 | 2260 | 1.20 | .05 | .85 | -1.1 | .85 | -.8 | .57 | 1.24 | R5 |
| 44 | 849 | 2260 | 1.05 | .05 | .93 | -.5 | .91 | -.5 | .50 | 1.13 | R44 |
| 72 | 867 | 2260 | 1.00 | .05 | 1.05 | .4 | 1.07 | .4 | .41 | .90 | R72 |
| 4 | 880 | 2260 | .97 | .05 | 1.14 | 1.0 | 1.22 | 1.2 | .35 | .69 | R4 |
| 54 | 883 | 2260 | .96 | .05 | .97 | -.2 | .97 | -.2 | .48 | 1.06 | R54 |
| 62 | 902 | 2260 | .92 | .05 | 1.31 | 2.1 | 1.45 | 2.4 | .22 | .34 | R62 |
| 34 | 905 | 2260 | .91 | .05 | 1.28 | 1.9 | 1.33 | 1.8 | .22 | .43 | R34 |
| 75 | 916 | 2260 | .89 | .05 | .95 | -.4 | .92 | -.5 | .50 | 1.11 | R75 |
| 99 | 952 | 2260 | .80 | .05 | .87 | -1.1 | .85 | -1.0 | .56 | 1.27 | R99 |
| 80 | 974 | 2260 | .75 | .05 | 1.05 | .4 | 1.06 | .4 | .42 | .89 | R80 |
| 94 | 989 | 2260 | .71 | .05 | .95 | -.4 | .96 | -.3 | .49 | 1.10 | R94 |
| 61 | 991 | 2260 | .71 | .05 | 1.07 | .6 | 1.11 | .7 | .37 | .81 | R61 |
| 56 | 1029 | 2260 | .62 | .05 | 1.08 | .6 | 1.10 | .6 | .38 | .81 | R56 |
| 97 | 1039 | 2260 | .60 | .05 | 1.01 | .1 | 1.02 | .1 | .45 | .97 | R97 |
| 52 | 1041 | 2260 | .59 | .05 | 1.34 | 2.5 | 1.45 | 2.5 | .18 | .19 | R52 |
| 88 | 1063 | 2260 | .54 | .05 | 1.05 | .4 | 1.08 | .5 | .40 | .86 | R88 |
| 76 | 1066 | 2260 | .54 | .05 | 1.34 | 2.6 | 1.41 | 2.3 | .17 | .17 | R76 |
| 87 | 1077 | 2260 | .51 | .05 | .88 | -1.1 | .86 | -.9 | .55 | 1.29 | R87 |
| 8 | 1087 | 2260 | .49 | .05 | 1.07 | .6 | 1.13 | .8 | .39 | .80 | R8 |
| 26 | 1093 | 2260 | .47 | .05 | .84 | -1.4 | .81 | -1.3 | .58 | 1.38 | R26 |
| 15 | 1094 | 2260 | .47 | .05 | .88 | -1.0 | .89 | -.8 | .55 | 1.27 | R15 |
| 33 | 1097 | 2260 | .47 | .05 | 1.09 | .7 | 1.04 | .3 | .38 | .81 | R33 |
| 43 | 1137 | 2260 | .37 | .05 | .93 | -.6 | .93 | -.5 | .50 | 1.16 | R43 |
| 77 | 1140 | 2260 | .37 | .05 | 1.08 | .6 | 1.14 | .8 | .38 | .79 | R77 |
| 55 | 1160 | 2260 | .32 | .05 | 1.15 | 1.2 | 1.13 | .8 | .32 | .64 | R55 |
| 78 | 1192 | 2260 | .25 | .05 | 1.17 | 1.4 | 1.23 | 1.3 | .30 | .53 | R78 |
| 59 | 1201 | 2260 | .23 | .05 | .77 | -2.2 | .70 | -2.0 | .64 | 1.60 | R59 |
| 6 | 1205 | 2260 | .22 | .05 | .91 | -.8 | .88 | -.7 | .51 | 1.23 | R6 |
| 68 | 1210 | 2260 | .21 | .05 | .87 | -1.2 | .82 | -1.2 | .55 | 1.35 | R68 |
| 74 | 1210 | 2260 | .21 | .05 | 1.04 | .3 | 1.02 | .1 | .41 | .92 | R74 |
| 71 | 1212 | 2260 | .21 | .05 | 1.14 | 1.2 | 1.24 | 1.3 | .32 | .59 | R71 |
| 95 | 1221 | 2260 | .19 | .05 | .88 | -1.2 | .80 | -1.3 | .55 | 1.35 | R95 |
| 69 | 1232 | 2260 | .16 | .05 | 1.19 | 1.6 | 1.29 | 1.6 | .28 | .46 | R69 |
| 64 | 1240 | 2260 | .14 | .05 | 1.03 | .3 | 1.02 | .1 | .41 | .91 | R64 |
| 81 | 1242 | 2260 | .14 | .05 | 1.20 | 1.7 | 1.23 | 1.2 | .27 | .48 | R81 |
| 30 | 1244 | 2260 | .13 | .05 | .79 | -2.1 | .71 | -1.9 | .62 | 1.58 | R30 |
| 32 | 1251 | 2260 | .12 | .05 | .85 | -1.5 | .79 | -1.3 | .57 | 1.41 | R32 |
| 31 | 1267 | 2260 | .08 | .05 | .91 | -.8 | .84 | -1.0 | .51 | 1.26 | R31 |
| 22 | 1288 | 2260 | .03 | .05 | .97 | -.3 | .94 | -.3 | .46 | 1.08 | R22 |
| 19 | 1294 | 2260 | .02 | .05 | .89 | -1.1 | .85 | -.9 | .53 | 1.30 | R19 |
| 57 | 1295 | 2260 | .02 | .05 | .87 | -1.2 | .79 | -1.2 | .55 | 1.36 | R57 |
| 93 | 1302 | 2260 | .00 | .05 | .87 | -1.3 | .80 | -1.2 | .55 | 1.37 | R93 |
| 2 | 1316 | 2260 | -.03 | .05 | .94 | -.6 | .89 | -.6 | .48 | 1.18 | R2 |
| 13 | 1324 | 2260 | -.05 | .05 | .96 | -.4 | .92 | -.5 | .46 | 1.12 | R13 |
| 21 | 1328 | 2260 | -.06 | .05 | 1.12 | 1.1 | 1.16 | .8 | .32 | .67 | R21 |
| 98 | 1338 | 2260 | -.08 | .05 | .86 | -1.3 | .78 | -1.2 | .55 | 1.38 | R98 |
| 36 | 1342 | 2260 | -.09 | .05 | .96 | -.4 | .89 | -.6 | .46 | 1.13 | R36 |
| 58 | 1342 | 2260 | -.09 | .05 | 1.05 | .5 | .98 | -.1 | .38 | .89 | R58 |
| 23 | 1351 | 2260 | -.11 | .05 | .95 | -.5 | .93 | -.4 | .47 | 1.14 | R23 |
| 90 | 1355 | 2260 | -.12 | .05 | 1.00 | .0 | 1.03 | .2 | .42 | .98 | R90 |
| 73 | 1360 | 2260 | -.13 | .05 | 1.00 | .0 | 1.00 | .0 | .42 | 1.00 | R73 |
| 7 | 1372 | 2260 | -.16 | .05 | .88 | -1.2 | .78 | -1.2 | .53 | 1.35 | R7 |
| 37 | 1380 | 2260 | -.18 | .05 | .95 | -.5 | .87 | -.7 | .47 | 1.16 | R37 |
| 67 | 1382 | 2260 | -.18 | .05 | 1.04 | .3 | 1.14 | .7 | .38 | .86 | R67 |
| 89 | 1404 | 2260 | -.23 | .05 | 1.09 | .8 | 1.11 | .5 | .34 | .77 | R89 |
| 27 | 1433 | 2260 | -.30 | .05 | .83 | -1.7 | .72 | -1.4 | .57 | 1.46 | R27 |
| 70 | 1454 | 2260 | -.35 | .05 | 1.12 | 1.0 | 1.32 | 1.3 | .29 | .63 | R70 |
| 65 | 1465 | 2260 | -.37 | .05 | .88 | -1.1 | .80 | -1.0 | .52 | 1.31 | R65 |
| 3 | 1486 | 2260 | -.42 | .05 | .97 | -.3 | .92 | -.4 | .43 | 1.08 | R3 |
| 28 | 1492 | 2260 | -.44 | .05 | 1.02 | .2 | 1.06 | .3 | .38 | .93 | R28 |
| 35 | 1494 | 2260 | -.44 | .05 | 1.09 | .7 | 1.06 | .3 | .32 | .80 | R35 |
| 11 | 1512 | 2260 | -.49 | .05 | .86 | -1.2 | .77 | -1.1 | .52 | 1.34 | R11 |
| 14 | 1522 | 2260 | -.51 | .05 | 1.14 | 1.1 | 1.24 | .9 | .26 | .66 | R14 |
| 53 | 1523 | 2260 | -.51 | .05 | 1.10 | .8 | 1.07 | .3 | .31 | .78 | R53 |
| 92 | 1538 | 2260 | -.55 | .05 | .80 | -1.8 | .67 | -1.5 | .58 | 1.47 | R92 |
| 60 | 1573 | 2260 | -.64 | .05 | .90 | -.9 | .83 | -.7 | .48 | 1.22 | R60 |
| 24 | 1600 | 2260 | -.71 | .05 | .87 | -1.1 | .77 | -.9 | .50 | 1.29 | R24 |
| 40 | 1606 | 2260 | -.72 | .05 | .93 | -.6 | .80 | -.8 | .45 | 1.17 | R40 |
| 16 | 1617 | 2260 | -.75 | .05 | 1.30 | 2.1 | 1.75 | 2.2 | .08 | .31 | R16 |
| 63 | 1628 | 2260 | -.78 | .05 | 1.10 | .7 | 1.00 | .0 | .30 | .85 | R63 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 49 | 1641 | 2260 | -.81 | .05 | .90 | -.8 | .78 | -.8 | .46 | 1.21 | R49 |
| 84 | 1655 | 2260 | -.85 | .05 | .94 | -.4 | .84 | -.6 | .42 | 1.12 | R84 |
| 18 | 1668 | 2260 | -.88 | .05 | .91 | -.7 | .80 | -.7 | .46 | 1.19 | R18 |
| 29 | 1681 | 2260 | -.92 | .05 | .96 | -.3 | .84 | -.6 | .41 | 1.10 | R29 |
| 48 | 1714 | 2260 | -1.01 | .05 | .92 | -.6 | .79 | -.7 | .43 | 1.15 | R48 |
| 1 | 1719 | 2260 | -1.02 | .05 | 1.00 | .0 | 1.08 | .2 | .34 | .98 | R1 |
| 50 | 1746 | 2260 | -1.10 | .05 | .92 | -.5 | .91 | -.3 | .40 | 1.10 | R50 |
| 51 | 1816 | 2260 | -1.32 | .06 | 1.04 | .2 | 1.23 | .6 | .26 | .90 | R51 |
| 39 | 1818 | 2260 | -1.32 | .06 | .97 | -.2 | .76 | -.7 | .37 | 1.07 | R39 |
| 82 | 1836 | 2260 | -1.38 | .06 | .90 | -.6 | .70 | -.8 | .42 | 1.15 | R82 |
| 17 | 1862 | 2260 | -1.47 | .06 | .95 | -.3 | .92 | -.2 | .35 | 1.05 | R17 |
| 85 | 1883 | 2260 | -1.54 | .06 | .93 | -.4 | .73 | -.7 | .38 | 1.11 | R85 |
| 41 | 1891 | 2260 | -1.57 | .06 | 1.06 | .3 | **1.38** | .8 | .21 | .89 | R41 |
| 47 | 1900 | 2260 | -1.60 | .06 | .91 | -.4 | .78 | -.5 | .38 | 1.10 | R47 |
| 46 | 1906 | 2260 | -1.63 | .06 | .87 | -.6 | .64 | -.9 | .42 | 1.16 | R46 |
| 45 | 1922 | 2260 | -1.69 | .06 | .85 | -.7 | .58 | -1.1 | .44 | 1.19 | R45 |
| 10 | 1994 | 2260 | -1.99 | .07 | .93 | -.3 | .70 | -.6 | .34 | 1.08 | R10 |
| 9 | 1997 | 2260 | -2.00 | .07 | 1.10 | .4 | **2.42** | **1.8** | .08 | .81 | R9 |
| 42 | 1999 | 2260 | -2.01 | .07 | .98 | -.1 | 1.19 | .3 | .26 | 1.00 | R42 |
| MEAN | 1289. | 2260. | .00 | .05 | 1.00 | -.1 | 1.00 | .0 | | | |
| S.D. | 361. | 0. | .90 | .00 | .13 | 1.0 | .26 | 1.0 | | | |

**Paso 5.** Revisión de los resultados observados para los sustentantes se presentan en dos segmentos por razones de espacio: los primeros 200 con medida más alta y los 200 con medida más baja al final:

```
TABLE 17.1 Analisis Defintivo EXEDII                ZOU765ws.txt Sep  4 12:12
2007
INPUT: 2260 PERSONS, 100 ITEMS  MEASURED: 2260 PERSONS, 100 ITEMS, 2 CATS
3.37
-------------------------------------------------------------------------------
-
PERSON: REAL SEP.: 4.37  REL.: .95 ... ITEM: REAL SEP.: 17.25  REL.:
1.00
```

PERSON STATISTICS: MEASURE ORDER

| ENTRY NUMBER | RAW SCORE | COUNT | MEASURE | ERROR | INFIT MNSQ | INFIT ZEMP | OUTFIT MNSQ | OUTFIT ZEMP | PTBIS CORR. | PERSON |
|---|---|---|---|---|---|---|---|---|---|---|
| 1303 | 99 | 100 | 4.96 | 1.01 | .99 | .0 | .43 | -.5 | .00 | 00253374 |
| 1311 | 99 | 100 | 4.96 | 1.01 | 1.00 | .0 | .58 | -.3 | .00 | 00252879 |
| 1381 | 99 | 100 | 4.96 | 1.01 | .99 | .0 | .43 | -.5 | .00 | 00253374 |
| 1594 | 99 | 100 | 4.96 | 1.01 | 1.01 | .0 | .99 | .0 | .00 | 00245437 |
| 1663 | 99 | 100 | 4.96 | 1.01 | 1.01 | .0 | .79 | -.1 | .00 | 00256036 |
| 563 | 98 | 100 | 4.24 | .72 | 1.03 | .1 | 1.11 | .1 | .00 | 00244379 |
| 952 | 98 | 100 | 4.24 | .72 | 1.00 | .0 | .61 | -.4 | .00 | 00306238 |
| 1515 | 98 | 100 | 4.24 | .72 | 1.04 | .1 | **1.98** | .7 | .00 | 00253987 |
| 273 | 97 | 100 | 3.82 | .59 | 1.00 | .0 | .71 | -.4 | .00 | 00250196 |
| 994 | 97 | 100 | 3.82 | .59 | .94 | -.1 | .52 | -.7 | .00 | 00305932 |
| 1126 | 97 | 100 | 3.82 | .59 | .94 | -.1 | .52 | -.7 | .00 | 00305932 |
| 1569 | 97 | 100 | 3.82 | .59 | 1.01 | .0 | .82 | -.2 | .00 | 00244437 |
| 2246 | 97 | 100 | 3.82 | .59 | .91 | -.2 | .36 | -1.0 | .00 | 00306387 |
| 190 | 96 | 100 | 3.52 | .52 | .94 | -.2 | .46 | -.9 | .00 | 00246776 |
| 583 | 96 | 100 | 3.52 | .52 | .99 | .0 | .66 | -.5 | .00 | 00249534 |
| 1392 | 96 | 100 | 3.52 | .52 | 1.02 | .0 | 1.05 | .1 | .00 | 00245383 |
| 1442 | 96 | 100 | 3.52 | .52 | .98 | .0 | .64 | -.6 | .00 | 00163512 |
| 1462 | 96 | 100 | 3.52 | .52 | .98 | .0 | .64 | -.6 | .00 | 00163512 |
| 1554 | 96 | 100 | 3.52 | .52 | 1.01 | .0 | .64 | -.6 | .00 | 00160076 |
| 1708 | 96 | 100 | 3.52 | .52 | 1.06 | .1 | 1.25 | .3 | .00 | 00245897 |
| 272 | 95 | 100 | 3.27 | .47 | 1.01 | .0 | 1.24 | .3 | .00 | 00250092 |
| 495 | 95 | 100 | 3.27 | .47 | .96 | -.1 | .66 | -.6 | .00 | 00159529 |
| 729 | 95 | 100 | 3.27 | .47 | 1.00 | .0 | .65 | -.6 | .00 | 00157109 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 822 | 95 | 100 | 3.27 | .47 | 1.04 | .1 | 1.20 | .3 | .00 | 00142153 |
| 824 | 95 | 100 | 3.27 | .47 | .99 | .0 | .87 | -.2 | .00 | 00154045 |
| 1080 | 95 | 100 | 3.27 | .47 | .86 | -.4 | .37 | -1.3 | .00 | 00306314 |
| 1144 | 95 | 100 | 3.27 | .47 | 1.03 | .1 | **2.61** | 1.7 | .00 | 00247054 |
| 1302 | 95 | 100 | 3.27 | .47 | .95 | -.2 | .52 | -.9 | .00 | 00243711 |
| 1325 | 95 | 100 | 3.27 | .47 | .94 | -.2 | .74 | -.4 | .00 | 00248351 |
| 1367 | 95 | 100 | 3.27 | .47 | .95 | -.2 | .52 | -.9 | .00 | 00243711 |
| 1648 | 95 | 100 | 3.27 | .47 | .95 | -.1 | .50 | -.9 | .00 | 00253135 |
| 1753 | 95 | 100 | 3.27 | .47 | 1.03 | .1 | 1.20 | .3 | .00 | 00158937 |
| 1889 | 95 | 100 | 3.27 | .47 | 1.01 | .0 | .72 | -.5 | .00 | 00245894 |
| 2118 | 95 | 100 | 3.27 | .47 | 1.01 | .0 | 1.78 | 1.0 | .00 | 00310768 |
| 2127 | 95 | 100 | 3.27 | .47 | .96 | -.1 | .61 | -.7 | .00 | 00243716 |
| 2259 | 95 | 100 | 3.27 | .47 | 1.00 | .0 | 1.01 | .0 | .00 | 00309023 |
| 44 | 94 | 100 | 3.07 | .43 | .96 | -.1 | **2.19** | 1.5 | .00 | 00242342 |
| 327 | 94 | 100 | 3.07 | .43 | 1.00 | .0 | .73 | -.5 | .00 | 00251219 |
| 548 | 94 | 100 | 3.07 | .43 | .94 | -.2 | .71 | -.6 | .00 | 00241103 |
| 650 | 94 | 100 | 3.07 | .43 | .89 | -.4 | .52 | -1.0 | .00 | 00157773 |
| 857 | 94 | 100 | 3.07 | .43 | 1.04 | .1 | **1.56** | .8 | .00 | 00307543 |
| 1000 | 94 | 100 | 3.07 | .43 | 1.09 | .3 | **1.54** | .8 | .00 | 00306432 |
| 1246 | 94 | 100 | 3.07 | .43 | .99 | .0 | **2.18** | 1.5 | .00 | 00157528 |
| 1436 | 94 | 100 | 3.07 | .43 | 1.09 | .3 | **3.23** | 2.5 | .00 | 00158816 |
| 1447 | 94 | 100 | 3.07 | .43 | 1.02 | .1 | **2.26** | 1.6 | .00 | 00163581 |
| 1456 | 94 | 100 | 3.07 | .43 | 1.09 | .3 | **3.23** | **2.5** | .00 | 00158816 |
| 1467 | 94 | 100 | 3.07 | .43 | 1.02 | .1 | **2.26** | 1.6 | .00 | 00163581 |
| 1526 | 94 | 100 | 3.07 | .43 | .95 | -.1 | .68 | -.6 | .00 | 00245514 |
| 1585 | 94 | 100 | 3.07 | .43 | 1.01 | .0 | .86 | -.3 | .00 | 00256570 |
| 1768 | 94 | 100 | 3.07 | .43 | .98 | -.1 | .87 | -.2 | .00 | 00164409 |
| 1896 | 94 | 100 | 3.07 | .43 | .98 | -.1 | .64 | -.7 | .00 | 00301637 |
| 1902 | 94 | 100 | 3.07 | .43 | 1.04 | .1 | 1.21 | .3 | .00 | 00306070 |
| 1905 | 94 | 100 | 3.07 | .43 | .97 | -.1 | .82 | -.3 | .00 | 00306208 |
| 2002 | 94 | 100 | 3.07 | .43 | .97 | -.1 | .76 | -.4 | .00 | 00308644 |
| 2152 | 94 | 100 | 3.07 | .43 | .94 | -.2 | .53 | -1.0 | .00 | 00245761 |
| 2188 | 94 | 100 | 3.07 | .43 | .97 | -.1 | .89 | -.2 | .00 | 00161910 |
| 2214 | 94 | 100 | 3.07 | .43 | .97 | -.1 | .89 | -.2 | .00 | 00161910 |
| 58 | 93 | 100 | 2.90 | .40 | 1.04 | .1 | **1.98** | 1.4 | .00 | 00246895 |
| 169 | 93 | 100 | 2.90 | .40 | .90 | -.3 | .65 | -.7 | .00 | 00160767 |
| 232 | 93 | 100 | 2.90 | .40 | 1.05 | .2 | 1.01 | .0 | .00 | 00251114 |
| 370 | 93 | 100 | 2.90 | .40 | .99 | .0 | **2.03** | 1.5 | .00 | 00242624 |
| 390 | 93 | 100 | 2.90 | .40 | .93 | -.3 | .56 | -1.0 | .00 | 00245907 |
| 525 | 93 | 100 | 2.90 | .40 | .91 | -.3 | .60 | -.9 | .00 | 00243843 |
| 701 | 93 | 100 | 2.90 | .40 | .93 | -.2 | .59 | -.9 | .00 | 00158149 |
| 793 | 93 | 100 | 2.90 | .40 | .99 | .0 | 1.00 | .0 | .00 | 00171156 |
| 862 | 93 | 100 | 2.90 | .40 | .98 | -.1 | .80 | -.4 | .00 | 00236413 |
| 1007 | 93 | 100 | 2.90 | .40 | 1.06 | .2 | 1.05 | .1 | .00 | 00235056 |
| 1286 | 93 | 100 | 2.90 | .40 | 1.01 | .0 | .82 | -.4 | .00 | 00245592 |
| 1402 | 93 | 100 | 2.90 | .40 | .97 | -.1 | .77 | -.5 | .00 | 00152464 |
| 1424 | 93 | 100 | 2.90 | .40 | .95 | -.2 | .85 | -.3 | .00 | 00250521 |
| 1581 | 93 | 100 | 2.90 | .40 | .95 | -.2 | .60 | -.9 | .00 | 00252357 |
| 1593 | 93 | 100 | 2.90 | .40 | .97 | -.1 | .84 | -.3 | .00 | 00244288 |
| 1596 | 93 | 100 | 2.90 | .40 | 1.02 | .1 | .86 | -.3 | .00 | 00245911 |
| 1615 | 93 | 100 | 2.90 | .40 | 1.02 | .1 | 1.14 | .3 | .00 | 00167748 |
| 1646 | 93 | 100 | 2.90 | .40 | .93 | -.2 | .55 | -1.0 | .00 | 00252913 |
| 1909 | 93 | 100 | 2.90 | .40 | .99 | .0 | .69 | -.7 | .00 | 00307232 |
| 2238 | 93 | 100 | 2.90 | .40 | 1.00 | .0 | **1.58** | .9 | .00 | 00305697 |
| 59 | 92 | 100 | 2.75 | .38 | .96 | -.2 | .64 | -.9 | .00 | 00246992 |
| 134 | 92 | 100 | 2.75 | .38 | .95 | -.2 | **1.73** | 1.2 | .00 | 00146825 |
| 167 | 92 | 100 | 2.75 | .38 | 1.05 | .2 | **1.88** | 1.4 | .00 | 00160680 |
| 191 | 92 | 100 | 2.75 | .38 | 1.04 | .1 | 1.11 | .2 | .00 | 00142519 |
| 577 | 92 | 100 | 2.75 | .38 | .95 | -.2 | .68 | -.7 | .00 | 00247670 |
| 653 | 92 | 100 | 2.75 | .38 | .93 | -.3 | **1.60** | 1.0 | .00 | 00167439 |
| 856 | 92 | 100 | 2.75 | .38 | .95 | -.2 | .65 | -.8 | .00 | 00306975 |
| 869 | 92 | 100 | 2.75 | .38 | 1.02 | .1 | .80 | -.4 | .00 | 00238971 |
| 937 | 92 | 100 | 2.75 | .38 | .95 | -.2 | .80 | -.4 | .00 | 00304656 |
| 1212 | 92 | 100 | 2.75 | .38 | 1.07 | .2 | .95 | -.1 | .00 | 00238901 |
| 1238 | 92 | 100 | 2.75 | .38 | 1.06 | .2 | **1.96** | 1.5 | .00 | 00152386 |
| 1239 | 92 | 100 | 2.75 | .38 | 1.07 | .3 | **1.39** | .7 | .00 | 00152699 |
| 1320 | 92 | 100 | 2.75 | .38 | .95 | -.2 | .60 | -.9 | .00 | 00245341 |
| 1423 | 92 | 100 | 2.75 | .38 | .99 | -.1 | .97 | -.1 | .00 | 00250474 |
| 1426 | 92 | 100 | 2.75 | .38 | 1.05 | .2 | **1.49** | .9 | .00 | 00250816 |
| 1525 | 92 | 100 | 2.75 | .38 | .95 | -.2 | .76 | -.5 | .00 | 00244062 |
| 1592 | 92 | 100 | 2.75 | .38 | .93 | -.3 | .66 | -.8 | .00 | 00243560 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1607 | 92 | 100 | 2.75 | .38 | .94 | -.2 | .57 | -1.0 | .00 | 00159980 |
| 1644 | 92 | 100 | 2.75 | .38 | .97 | -.1 | .70 | -.7 | .00 | 00251224 |
| 1661 | 92 | 100 | 2.75 | .38 | .95 | -.2 | .87 | -.3 | .00 | 00251576 |
| 1827 | 92 | 100 | 2.75 | .38 | 1.04 | .2 | .93 | -.1 | .00 | 00164574 |
| 1972 | 92 | 100 | 2.75 | .38 | 1.03 | .1 | .98 | -.1 | .00 | 00251126 |
| 2083 | 92 | 100 | 2.75 | .38 | .94 | -.2 | .63 | -.9 | .00 | 00165752 |
| 2141 | 92 | 100 | 2.75 | .38 | .96 | -.1 | .72 | -.6 | .00 | 00308925 |
| 2243 | 92 | 100 | 2.75 | .38 | .94 | -.2 | .65 | -.8 | .00 | 00306171 |
| 2257 | 92 | 100 | 2.75 | .38 | .93 | -.3 | .77 | -.5 | .00 | 00308429 |
| 168 | 91 | 100 | 2.61 | .36 | 1.02 | .1 | 1.05 | .1 | .00 | 00160689 |
| 231 | 91 | 100 | 2.61 | .36 | .89 | -.4 | .60 | -1.0 | .00 | 00250714 |
| 243 | 91 | 100 | 2.61 | .36 | 1.02 | .1 | .86 | -.3 | .00 | 00169499 |
| 263 | 91 | 100 | 2.61 | .36 | .92 | -.3 | .62 | -1.0 | .00 | 00247276 |
| 302 | 91 | 100 | 2.61 | .36 | .99 | .0 | 1.00 | .0 | .00 | 00241107 |
| 373 | 91 | 100 | 2.61 | .36 | .94 | -.3 | .68 | -.8 | .00 | 00242862 |
| 397 | 91 | 100 | 2.61 | .36 | .95 | -.2 | .62 | -1.0 | .00 | 00250102 |
| 410 | 91 | 100 | 2.61 | .36 | 1.06 | .2 | 1.03 | .1 | .00 | 00247196 |
| 421 | 91 | 100 | 2.61 | .36 | 1.04 | .2 | 1.15 | .3 | .00 | 10124740 |
| 475 | 91 | 100 | 2.61 | .36 | .94 | -.2 | .85 | -.4 | .00 | 00137768 |
| 578 | 91 | 100 | 2.61 | .36 | .96 | -.1 | .79 | -.5 | .00 | 00247718 |
| 618 | 91 | 100 | 2.61 | .36 | .93 | -.3 | .68 | -.8 | .00 | 00160621 |
| 841 | 91 | 100 | 2.61 | .36 | .93 | -.3 | .69 | -.8 | .00 | 00303715 |
| 906 | 91 | 100 | 2.61 | .36 | .85 | -.6 | .50 | -1.4 | .00 | 00248442 |
| 911 | 91 | 100 | 2.61 | .36 | .96 | -.2 | **1.55** | 1.0 | .00 | 00250756 |
| 914 | 91 | 100 | 2.61 | .36 | 1.01 | .0 | 1.04 | .1 | .00 | 00251215 |
| 1082 | 91 | 100 | 2.61 | .36 | 1.00 | .0 | .95 | -.1 | .00 | 00306974 |
| 1103 | 91 | 100 | 2.61 | .36 | .90 | -.4 | .77 | -.6 | .00 | 00247187 |
| 1109 | 91 | 100 | 2.61 | .36 | .91 | -.3 | .67 | -.8 | .00 | 00247369 |
| 1210 | 91 | 100 | 2.61 | .36 | .96 | -.2 | .86 | -.3 | .00 | 00237551 |
| 1219 | 91 | 100 | 2.61 | .36 | 1.00 | .0 | .94 | -.1 | .00 | 00243587 |
| 1293 | 91 | 100 | 2.61 | .36 | .89 | -.5 | **1.44** | .8 | .00 | 00247031 |
| 1336 | 91 | 100 | 2.61 | .36 | .97 | -.1 | .85 | -.3 | .00 | 00158751 |
| 1641 | 91 | 100 | 2.61 | .36 | .98 | -.1 | .99 | .0 | .00 | 00249801 |
| 1797 | 91 | 100 | 2.61 | .36 | 1.01 | .0 | **1.35** | .7 | .00 | 00302168 |
| 1870 | 91 | 100 | 2.61 | .36 | .84 | -.7 | .51 | -1.3 | .00 | 00255113 |
| 1922 | 91 | 100 | 2.61 | .36 | .97 | -.1 | .66 | -.9 | .00 | 00157268 |
| 1941 | 91 | 100 | 2.61 | .36 | .98 | -.1 | **1.62** | 1.1 | .00 | 00241953 |
| 1985 | 91 | 100 | 2.61 | .36 | 1.07 | .3 | 1.26 | .5 | .00 | 00176026 |
| 2003 | 91 | 100 | 2.61 | .36 | .88 | -.5 | .59 | -1.1 | .00 | 00309763 |
| 1 | 90 | 100 | 2.49 | .34 | .91 | -.4 | .57 | -1.2 | .00 | 00160335 |
| 70 | 90 | 100 | 2.49 | .34 | 1.15 | .6 | **2.34** | **2.3** | .00 | 00256963 |
| 124 | 90 | 100 | 2.49 | .34 | .98 | -.1 | .75 | -.6 | .00 | 00164233 |
| 223 | 90 | 100 | 2.49 | .34 | .95 | -.2 | 1.03 | .1 | .00 | 00243597 |
| 360 | 90 | 100 | 2.49 | .34 | .86 | -.6 | .63 | -1.0 | .00 | 00240496 |
| 381 | 90 | 100 | 2.49 | .34 | .92 | -.3 | .76 | -.6 | .00 | 00243933 |
| 445 | 90 | 100 | 2.49 | .34 | .93 | -.3 | .90 | -.2 | .00 | 00245424 |
| 529 | 90 | 100 | 2.49 | .34 | .99 | -.1 | .83 | -.4 | .00 | 00244833 |
| 871 | 90 | 100 | 2.49 | .34 | .92 | -.3 | .65 | -.9 | .00 | 00239705 |
| 904 | 90 | 100 | 2.49 | .34 | 1.04 | .2 | **2.06** | 1.9 | .00 | 00247898 |
| 948 | 90 | 100 | 2.49 | .34 | 1.01 | .0 | **1.57** | 1.1 | .00 | 00305532 |
| 975 | 90 | 100 | 2.49 | .34 | 1.09 | .4 | 1.01 | .0 | .00 | 00248025 |
| 1020 | 90 | 100 | 2.49 | .34 | .91 | -.4 | 1.38 | .8 | .00 | 00156637 |
| 1296 | 90 | 100 | 2.49 | .34 | 1.01 | .1 | 1.02 | .1 | .00 | 00248043 |
| 1430 | 90 | 100 | 2.49 | .34 | .99 | -.1 | 1.07 | .2 | .00 | 00251279 |
| 1435 | 90 | 100 | 2.49 | .34 | 1.08 | .3 | 1.02 | .0 | .00 | 00150210 |
| 1445 | 90 | 100 | 2.49 | .34 | .99 | .0 | .88 | -.3 | .00 | 00163564 |
| 1455 | 90 | 100 | 2.49 | .34 | 1.08 | .3 | 1.02 | .0 | .00 | 00150210 |
| 1465 | 90 | 100 | 2.49 | .34 | .99 | .0 | .88 | -.3 | .00 | 00163564 |
| 1605 | 90 | 100 | 2.49 | .34 | .96 | -.2 | .85 | -.4 | .00 | 00153040 |
| 1618 | 90 | 100 | 2.49 | .34 | .97 | -.1 | 1.22 | .5 | .00 | 00153899 |
| 1639 | 90 | 100 | 2.49 | .34 | .99 | -.1 | .92 | -.2 | .00 | 00246019 |
| 1975 | 90 | 100 | 2.49 | .34 | 1.02 | .1 | .86 | -.4 | .00 | 00253591 |
| 2004 | 90 | 100 | 2.49 | .34 | .93 | -.3 | .74 | -.7 | .00 | 00309818 |
| 2037 | 90 | 100 | 2.49 | .34 | 1.07 | .3 | 1.04 | .1 | .00 | 00150996 |
| 91 | 89 | 100 | 2.38 | .33 | .98 | -.1 | .79 | -.6 | .00 | 00235341 |
| 178 | 89 | 100 | 2.38 | .33 | .97 | -.1 | .99 | .0 | .00 | 00161696 |
| 200 | 89 | 100 | 2.38 | .33 | .92 | -.4 | .65 | -1.0 | .00 | 00156082 |
| 424 | 89 | 100 | 2.38 | .33 | 1.05 | .2 | **1.80** | 1.6 | .00 | 10128434 |
| 430 | 89 | 100 | 2.38 | .33 | .92 | -.4 | **1.38** | .9 | .00 | 00240168 |
| 440 | 89 | 100 | 2.38 | .33 | 1.08 | .4 | **1.44** | 1.0 | .00 | 00244126 |
| 818 | 89 | 100 | 2.38 | .33 | 1.10 | .4 | .94 | -.1 | .00 | 00248830 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 849 | 89 | 100 | 2.38 | .33 | .92 | -.4 | .76 | -.7 | .00 | 00305297 |
| | 1202 | 89 | 100 | 2.38 | .33 | 1.03 | .1 | .95 | -.1 | .00 | 00164231 |
| | 1227 | 89 | 100 | 2.38 | .33 | .97 | -.1 | .81 | -.5 | .00 | 00247327 |
| | 1307 | 89 | 100 | 2.38 | .33 | 1.06 | .3 | **2.01** | 2.0 | .00 | 00248535 |
| | 1431 | 89 | 100 | 2.38 | .33 | .97 | -.1 | .80 | -.5 | .00 | 00264992 |
| | 1433 | 89 | 100 | 2.38 | .33 | 1.04 | .2 | .91 | -.2 | .00 | 00145297 |
| | 1453 | 89 | 100 | 2.38 | .33 | 1.04 | .2 | .91 | -.2 | .00 | 00145297 |
| | 1631 | 89 | 100 | 2.38 | .33 | .99 | -.1 | .78 | -.6 | .00 | 00165855 |
| | 1719 | 89 | 100 | 2.38 | .33 | .94 | -.3 | .64 | -1.1 | .00 | 00254672 |
| | 1795 | 89 | 100 | 2.38 | .33 | 1.04 | .2 | .85 | -.4 | .00 | 00254684 |
| | 1829 | 89 | 100 | 2.38 | .33 | .92 | -.4 | .65 | -1.0 | .00 | 00165444 |
| | 1831 | 89 | 100 | 2.38 | .33 | .98 | -.1 | .85 | -.4 | .00 | 00166029 |
| | 1888 | 89 | 100 | 2.38 | .33 | .95 | -.2 | .78 | -.6 | .00 | 00245582 |
| | 1927 | 89 | 100 | 2.38 | .33 | .98 | -.1 | 1.08 | .2 | .00 | 00165984 |
| | 1944 | 89 | 100 | 2.38 | .33 | .93 | -.3 | .70 | -.8 | .00 | 00243227 |
| | 2001 | 89 | 100 | 2.38 | .33 | .88 | -.5 | .60 | -1.2 | .00 | 00308007 |
| | 2082 | 89 | 100 | 2.38 | .33 | .96 | -.2 | .83 | -.4 | .00 | 00165738 |
| | 2253 | 89 | 100 | 2.38 | .33 | 1.09 | .4 | 1.26 | .6 | .00 | 00307503 |
| | 25 | 88 | 100 | 2.27 | .32 | .94 | -.3 | 1.05 | .1 | .00 | 00306159 |
| | 29 | 88 | 100 | 2.27 | .32 | .97 | -.1 | .82 | -.5 | .00 | 00310014 |
| | 53 | 88 | 100 | 2.27 | .32 | 1.01 | .0 | .89 | -.3 | .00 | 00245289 |
| | 125 | 88 | 100 | 2.27 | .32 | .85 | -.8 | .54 | -1.5 | .00 | 00164297 |
| | 187 | 88 | 100 | 2.27 | .32 | .88 | -.6 | .67 | -1.0 | .00 | 00163503 |
| | 276 | 88 | 100 | 2.27 | .32 | 1.07 | .3 | 1.02 | .1 | .00 | 00116609 |
| | 329 | 88 | 100 | 2.27 | .32 | .98 | -.1 | 1.01 | .0 | .00 | 00251418 |
| | 441 | 88 | 100 | 2.27 | .32 | .96 | -.2 | .98 | -.1 | .00 | 00244249 |
| | 526 | 88 | 100 | 2.27 | .32 | .95 | -.3 | .79 | -.6 | .00 | 00243905 |
| | 864 | 88 | 100 | 2.27 | .32 | .98 | -.1 | 1.21 | .5 | .00 | 00237359 |
| | 903 | 88 | 100 | 2.27 | .32 | .98 | -.1 | .95 | -.2 | .00 | 00247744 |
| | 934 | 88 | 100 | 2.27 | .32 | .95 | -.3 | .82 | -.5 | .00 | 00304523 |
| | 1199 | 88 | 100 | 2.27 | .32 | 1.01 | .1 | .93 | -.2 | .00 | 00163560 |
| | 1230 | 88 | 100 | 2.27 | .32 | 1.00 | .0 | .90 | -.3 | .00 | 00251484 |
| | 1413 | 88 | 100 | 2.27 | .32 | 1.01 | .1 | .90 | -.3 | .00 | 00244755 |
| | 1449 | 88 | 100 | 2.27 | .32 | .98 | -.1 | **1.31** | .8 | .00 | 00163807 |
| | 1469 | 88 | 100 | 2.27 | .32 | .98 | -.1 | 1.31 | .8 | .00 | 00163807 |

.....

.....

.....

|

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 157 | 30 | 100 | -.99 | .24 | 1.09 | .8 | 1.22 | 1.2 | .00 | 00158141 |
| | 236 | 30 | 100 | -.99 | .24 | 1.19 | 1.7 | **1.37** | 2.0 | .00 | 00149040 |
| | 333 | 30 | 100 | -.99 | .24 | .97 | -.3 | .91 | -.6 | .00 | 00130333 |
| | 356 | 30 | 100 | -.99 | .24 | 1.20 | 1.8 | **1.30** | 1.7 | .00 | 00231755 |
| | 442 | 30 | 100 | -.99 | .24 | 1.09 | .9 | **1.29** | 1.6 | .00 | 00244524 |
| | 544 | 30 | 100 | -.99 | .24 | 1.03 | .3 | 1.10 | .6 | .00 | 00237543 |
| | 597 | 30 | 100 | -.99 | .24 | 1.05 | .5 | 1.08 | .5 | .00 | 00152525 |
| | 611 | 30 | 100 | -.99 | .24 | 1.27 | 2.4 | 1.34 | 1.9 | .00 | 00158410 |
| | 690 | 30 | 100 | -.99 | .24 | 1.15 | 1.4 | 1.27 | 1.5 | .00 | 00152490 |
| | 697 | 30 | 100 | -.99 | .24 | .93 | -.7 | .87 | -.8 | .00 | 00157015 |
| | 705 | 30 | 100 | -.99 | .24 | 1.09 | .9 | 1.09 | .6 | .00 | 00158991 |
| | 719 | 30 | 100 | -.99 | .24 | 1.04 | .3 | 1.03 | .2 | .00 | 00149047 |
| | 742 | 30 | 100 | -.99 | .24 | .94 | -.6 | .93 | -.5 | .00 | 00146551 |
| | 755 | 30 | 100 | -.99 | .24 | 1.03 | .3 | 1.08 | .5 | .00 | 00157845 |
| | 760 | 30 | 100 | -.99 | .24 | .85 | -1.6 | .89 | -.7 | .00 | 00158935 |
| | 785 | 30 | 100 | -.99 | .24 | .98 | -.2 | .98 | -.1 | .00 | 00161296 |
| | 808 | 30 | 100 | -.99 | .24 | 1.02 | .2 | 1.16 | .9 | .00 | 00242364 |
| | 894 | 30 | 100 | -.99 | .24 | 1.01 | .1 | 1.00 | .0 | .00 | 00245736 |
| | 1056 | 30 | 100 | -.99 | .24 | 1.01 | .1 | 1.09 | .5 | .00 | 00154694 |
| | 1137 | 30 | 100 | -.99 | .24 | .96 | -.4 | .93 | -.5 | .00 | 00242287 |
| | 1150 | 30 | 100 | -.99 | .24 | **1.33** | **2.9** | **1.59** | **3.0** | .00 | 00235149 |
| | 1363 | 30 | 100 | -.99 | .24 | 1.05 | .5 | 1.07 | .4 | .00 | 00237367 |
| | 1523 | 30 | 100 | -.99 | .24 | 1.12 | 1.1 | 1.27 | 1.5 | .00 | 00243437 |
| | 1674 | 30 | 100 | -.99 | .24 | .91 | -.9 | .85 | -1.0 | .00 | 00161726 |
| | 1692 | 30 | 100 | -.99 | .24 | .91 | -.9 | .85 | -1.0 | .00 | 00161726 |
| | 1749 | 30 | 100 | -.99 | .24 | 1.11 | 1.0 | 1.21 | 1.2 | .00 | 00156696 |
| | 1838 | 30 | 100 | -.99 | .24 | 1.14 | 1.3 | **1.38** | **2.1** | .00 | 07508515 |
| | 1880 | 30 | 100 | -.99 | .24 | 1.18 | 1.7 | 1.29 | 1.6 | .00 | 00240043 |
| | 1987 | 30 | 100 | -.99 | .24 | .94 | -.6 | 1.03 | .2 | .00 | 00240043 |
| | 2019 | 30 | 100 | -.99 | .24 | .89 | -1.1 | .98 | -.1 | .00 | 00162454 |
| | 83 | 29 | 100 | -1.04 | .24 | 1.06 | .6 | 1.26 | 1.4 | .00 | 00156696 |
| | 156 | 29 | 100 | -1.04 | .24 | 1.08 | .8 | 1.04 | .2 | .00 | 00157494 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 197 | 29 | 100 | -1.04 | .24 | 1.27 | 2.4 | 1.38 | 2.0 | .00 | 00151653 |
| 218 | 29 | 100 | -1.04 | .24 | 1.14 | 1.3 | 1.13 | .8 | .00 | 00241879 |
| 299 | 29 | 100 | -1.04 | .24 | 1.06 | .5 | 1.08 | .5 | .00 | 00240417 |
| 465 | 29 | 100 | -1.04 | .24 | 1.06 | .6 | 1.08 | .5 | .00 | 00158371 |
| 512 | 29 | 100 | -1.04 | .24 | 1.05 | .5 | 1.07 | .4 | .00 | 00240019 |
| 574 | 29 | 100 | -1.04 | .24 | 1.05 | .4 | 1.05 | .3 | .00 | 00245736 |
| 613 | 29 | 100 | -1.04 | .24 | 1.15 | 1.4 | 1.05 | .3 | .00 | 00159485 |
| 730 | 29 | 100 | -1.04 | .24 | .96 | -.4 | 1.08 | .5 | .00 | 00158425 |
| 870 | 29 | 100 | -1.04 | .24 | 1.01 | .1 | 1.20 | 1.1 | .00 | 00239526 |
| 880 | 29 | 100 | -1.04 | .24 | 1.11 | 1.0 | 1.28 | 1.5 | .00 | 00242476 |
| 961 | 29 | 100 | -1.04 | .24 | 1.08 | .7 | 1.16 | .9 | .00 | 00242525 |
| 963 | 29 | 100 | -1.04 | .24 | 1.11 | 1.0 | 1.37 | 2.0 | .00 | 00242547 |
| 1088 | 29 | 100 | -1.04 | .24 | 1.08 | .7 | 1.16 | .9 | .00 | 00242525 |
| 1111 | 29 | 100 | -1.04 | .24 | 1.11 | 1.0 | 1.37 | 2.0 | .00 | 00242547 |
| 1156 | 29 | 100 | -1.04 | .24 | 1.07 | .7 | 1.15 | .8 | .00 | 00242667 |
| 1171 | 29 | 100 | -1.04 | .24 | .94 | -.6 | .87 | -.8 | .00 | 00306132 |
| 1265 | 29 | 100 | -1.04 | .24 | 1.11 | 1.0 | 1.31 | 1.6 | .00 | 00304090 |
| 1267 | 29 | 100 | -1.04 | .24 | 1.07 | .7 | 1.12 | .7 | .00 | 00304638 |
| 1270 | 29 | 100 | -1.04 | .24 | 1.06 | .6 | 1.13 | .7 | .00 | 00305319 |
| 1341 | 29 | 100 | -1.04 | .24 | .95 | -.4 | .97 | -.2 | .00 | 00161523 |
| 1481 | 29 | 100 | -1.04 | .24 | .96 | -.4 | 1.00 | .0 | .00 | 00154741 |
| 1799 | 29 | 100 | -1.04 | .24 | 1.22 | 2.0 | 1.30 | 1.6 | .00 | 00140878 |
| 1812 | 29 | 100 | -1.04 | .24 | 1.04 | .4 | 1.13 | .7 | .00 | 00155968 |
| 1825 | 29 | 100 | -1.04 | .24 | 1.27 | 2.4 | 1.41 | 2.1 | .00 | 00162603 |
| 1833 | 29 | 100 | -1.04 | .24 | 1.06 | .6 | 1.13 | .8 | .00 | 00166275 |
| 1875 | 29 | 100 | -1.04 | .24 | 1.04 | .3 | 1.05 | .3 | .00 | 00170287 |
| 1939 | 29 | 100 | -1.04 | .24 | 1.12 | 1.1 | 1.18 | 1.0 | .00 | 00237402 |
| 1979 | 29 | 100 | -1.04 | .24 | 1.12 | 1.1 | 1.25 | 1.4 | .00 | 00148634 |
| 2010 | 29 | 100 | -1.04 | .24 | .94 | -.6 | .96 | -.3 | .00 | 00153999 |
| 2015 | 29 | 100 | -1.04 | .24 | 1.09 | .8 | 1.21 | 1.2 | .00 | 00160382 |
| 2062 | 29 | 100 | -1.04 | .24 | 1.01 | .1 | 1.04 | .3 | .00 | 00155487 |
| 2111 | 29 | 100 | -1.04 | .24 | 1.16 | 1.5 | 1.30 | 1.6 | .00 | 00305631 |
| 6 | 28 | 100 | -1.10 | .24 | 1.17 | 1.5 | 1.47 | 2.3 | .00 | 00240066 |
| 116 | 28 | 100 | -1.10 | .24 | 1.33 | 2.8 | 1.54 | 2.6 | .00 | 00159522 |
| 306 | 28 | 100 | -1.10 | .24 | .96 | -.4 | .93 | -.4 | .00 | 00243305 |
| 447 | 28 | 100 | -1.10 | .24 | 1.02 | .2 | 1.09 | .5 | .00 | 00246111 |
| 508 | 28 | 100 | -1.10 | .24 | 1.06 | .5 | 1.10 | .5 | .00 | 00239109 |
| 518 | 28 | 100 | -1.10 | .24 | 1.08 | .7 | 1.19 | 1.0 | .00 | 00242599 |
| 606 | 28 | 100 | -1.10 | .24 | 1.00 | .0 | .97 | -.2 | .00 | 00156536 |
| 608 | 28 | 100 | -1.10 | .24 | .81 | -1.9 | .74 | -1.6 | .00 | 00156595 |
| 625 | 28 | 100 | -1.10 | .24 | .98 | -.2 | .96 | -.2 | .00 | 00132382 |
| 865 | 28 | 100 | -1.10 | .24 | 1.12 | 1.1 | 1.35 | 1.8 | .00 | 00238369 |
| 908 | 28 | 100 | -1.10 | .24 | 1.08 | .7 | 1.11 | .6 | .00 | 00248886 |
| 992 | 28 | 100 | -1.10 | .24 | .96 | -.4 | .91 | -.6 | .00 | 00305721 |
| 1135 | 28 | 100 | -1.10 | .24 | .93 | -.6 | .85 | -.9 | .00 | 00234970 |
| 1165 | 28 | 100 | -1.10 | .24 | 1.04 | .4 | 1.06 | .3 | .00 | 00304178 |
| 1181 | 28 | 100 | -1.10 | .24 | 1.20 | 1.7 | 1.53 | 2.6 | .00 | 00150108 |
| 1237 | 28 | 100 | -1.10 | .24 | 1.33 | 2.8 | 1.51 | 2.5 | .00 | 00151836 |
| 1366 | 28 | 100 | -1.10 | .24 | .96 | -.4 | .94 | -.4 | .00 | 00243404 |
| 1564 | 28 | 100 | -1.10 | .24 | 1.18 | 1.6 | 1.18 | 1.0 | .00 | 00237016 |
| 1757 | 28 | 100 | -1.10 | .24 | .96 | -.4 | .98 | -.1 | .00 | 00161012 |
| 2094 | 28 | 100 | -1.10 | .24 | .92 | -.7 | .78 | -1.3 | .00 | 00249301 |
| 2112 | 28 | 100 | -1.10 | .24 | 1.08 | .7 | 1.36 | 1.8 | .00 | 00305659 |
| 2169 | 28 | 100 | -1.10 | .24 | 1.05 | .5 | 1.15 | .8 | .00 | 00146682 |
| 2195 | 28 | 100 | -1.10 | .24 | 1.05 | .5 | 1.15 | .8 | .00 | 00146682 |
| 164 | 27 | 100 | -1.16 | .24 | 1.22 | 1.8 | 1.30 | 1.5 | .00 | 00159975 |
| 347 | 27 | 100 | -1.16 | .24 | 1.12 | 1.0 | 1.21 | 1.1 | .00 | 00158425 |
| 593 | 27 | 100 | -1.16 | .24 | 1.14 | 1.2 | 1.27 | 1.3 | .00 | 00146714 |
| 710 | 27 | 100 | -1.16 | .24 | .98 | -.1 | .91 | -.5 | .00 | 00161172 |
| 716 | 27 | 100 | -1.16 | .24 | 1.05 | .5 | .99 | -.1 | .00 | 00148587 |
| 778 | 27 | 100 | -1.16 | .24 | 1.19 | 1.6 | 1.27 | 1.4 | .00 | 00155313 |
| 977 | 27 | 100 | -1.16 | .24 | .95 | -.5 | .87 | -.7 | .00 | 00250232 |
| 1200 | 27 | 100 | -1.16 | .24 | 1.02 | .2 | 1.26 | 1.3 | .00 | 00163715 |
| 1300 | 27 | 100 | -1.16 | .24 | 1.03 | .3 | 1.08 | .4 | .00 | 00252483 |
| 1372 | 27 | 100 | -1.16 | .24 | 1.03 | .3 | 1.08 | .4 | .00 | 00248103 |
| 1884 | 27 | 100 | -1.16 | .24 | .98 | -.2 | 1.08 | .4 | .00 | 00242601 |
| 2008 | 27 | 100 | -1.16 | .24 | 1.15 | 1.3 | 1.36 | 1.8 | .00 | 00151653 |
| 2034 | 27 | 100 | -1.16 | .24 | 1.13 | 1.1 | 1.21 | 1.1 | .00 | 00170055 |
| 2115 | 27 | 100 | -1.16 | .24 | .93 | -.7 | .96 | -.2 | .00 | 00306953 |
| 2148 | 27 | 100 | -1.16 | .24 | .96 | -.3 | .91 | -.5 | .00 | 00243680 |
| 225 | 26 | 100 | -1.22 | .25 | 1.01 | .1 | 1.06 | .3 | .00 | 00244888 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 247 | 26 | 100 | -1.22 | .25 | 1.06 | .5 | 1.12 | .6 | .00 | 00236399 |
| 249 | 26 | 100 | -1.22 | .25 | 1.13 | 1.1 | **1.32** | 1.5 | .00 | 00236643 |
| 456 | 26 | 100 | -1.22 | .25 | **1.35** | **2.7** | **1.62** | **2.7** | .00 | 00151836 |
| 488 | 26 | 100 | -1.22 | .25 | .94 | -.6 | .90 | -.5 | .00 | 00156536 |
| 585 | 26 | 100 | -1.22 | .25 | 1.14 | 1.1 | 1.24 | 1.2 | .00 | 00132092 |
| 601 | 26 | 100 | -1.22 | .25 | 1.16 | 1.3 | 1.18 | .9 | .00 | 00153632 |
| 664 | 26 | 100 | -1.22 | .25 | 1.19 | 1.5 | **1.32** | 1.5 | .00 | 00240417 |
| 801 | 26 | 100 | -1.22 | .25 | 1.08 | .7 | 1.09 | .5 | .00 | 00237596 |
| 833 | 26 | 100 | -1.22 | .25 | **1.33** | **2.6** | **1.73** | **3.1** | .00 | 00166448 |
| 1050 | 26 | 100 | -1.22 | .25 | **1.26** | **2.1** | **1.52** | **2.3** | .00 | 00148641 |
| 1173 | 26 | 100 | -1.22 | .25 | 1.15 | 1.2 | 1.13 | .6 | .00 | 00306758 |
| 1242 | 26 | 100 | -1.22 | .25 | 1.20 | 1.6 | **1.38** | 1.8 | .00 | 00155313 |
| 1315 | 26 | 100 | -1.22 | .25 | 1.05 | .5 | 1.04 | .2 | .00 | 00146754 |
| 1598 | 26 | 100 | -1.22 | .25 | 1.02 | .1 | 1.09 | .5 | .00 | 00247854 |
| 1635 | 26 | 100 | -1.22 | .25 | 1.15 | 1.2 | **1.48** | **2.2** | .00 | 00233380 |
| 1754 | 26 | 100 | -1.22 | .25 | 1.01 | .1 | 1.11 | .6 | .00 | 00159874 |
| 1798 | 26 | 100 | -1.22 | .25 | 1.15 | 1.2 | **1.34** | 1.6 | .00 | 00128036 |
| 1862 | 26 | 100 | -1.22 | .25 | 1.03 | .3 | .99 | -.1 | .00 | 00246385 |
| 2005 | 26 | 100 | -1.22 | .25 | 1.13 | 1.1 | 1.25 | 1.2 | .00 | 00144912 |
| 2030 | 26 | 100 | -1.22 | .25 | 1.14 | 1.2 | **1.36** | 1.7 | .00 | 00166011 |
| 2045 | 26 | 100 | -1.22 | .25 | .89 | -1.0 | .91 | -.5 | .00 | 00161728 |
| 2166 | 26 | 100 | -1.22 | .25 | 1.05 | .5 | 1.01 | .1 | .00 | 00255532 |
| 2228 | 26 | 100 | -1.22 | .25 | 1.25 | **2.0** | **1.41** | **1.9** | .00 | 00303935 |
| 219 | 25 | 100 | -1.28 | .25 | 1.05 | .4 | 1.26 | 1.2 | .00 | 00242516 |
| 233 | 25 | 100 | -1.28 | .25 | 1.06 | .5 | 1.15 | .7 | .00 | 00132092 |
| 254 | 25 | 100 | -1.28 | .25 | .95 | -.4 | .92 | -.4 | .00 | 00243277 |
| 294 | 25 | 100 | -1.28 | .25 | .87 | -1.2 | .75 | -1.4 | .00 | 00163749 |
| 338 | 25 | 100 | -1.28 | .25 | 1.09 | .7 | 1.16 | .8 | .00 | 00149047 |
| 644 | 25 | 100 | -1.28 | .25 | 1.04 | .4 | 1.14 | .7 | .00 | 07000453 |
| 654 | 25 | 100 | -1.28 | .25 | 1.09 | .8 | 1.25 | 1.1 | .00 | 00170152 |
| 714 | 25 | 100 | -1.28 | .25 | 1.19 | 1.5 | **1.60** | **2.5** | .00 | 00144094 |
| 1062 | 25 | 100 | -1.28 | .25 | .98 | -.1 | .89 | -.6 | .00 | 00159485 |
| 1076 | 25 | 100 | -1.28 | .25 | 1.05 | .4 | 1.20 | 1.0 | .00 | 00304638 |
| 1157 | 25 | 100 | -1.28 | .25 | 1.11 | .9 | 1.23 | 1.1 | .00 | 00242679 |
| 1259 | 25 | 100 | -1.28 | .25 | 1.05 | .4 | .97 | -.1 | .00 | 00242539 |
| 1339 | 25 | 100 | -1.28 | .25 | 1.09 | .8 | 1.06 | .3 | .00 | 00160953 |
| 1745 | 25 | 100 | -1.28 | .25 | 1.09 | .8 | 1.06 | .3 | .00 | 00254827 |
| 1885 | 25 | 100 | -1.28 | .25 | 1.11 | .9 | **1.35** | 1.6 | .00 | 00242679 |
| 1984 | 25 | 100 | -1.28 | .25 | 1.12 | .9 | 1.22 | 1.0 | .00 | 00163705 |
| 2151 | 25 | 100 | -1.28 | .25 | 1.08 | .7 | 1.03 | .1 | .00 | 00245561 |
| 8 | 24 | 100 | -1.34 | .25 | **1.31** | **2.3** | **1.64** | **2.6** | .00 | 00242493 |
| 431 | 24 | 100 | -1.34 | .25 | 1.08 | .7 | 1.09 | .4 | .00 | 00240441 |
| 484 | 24 | 100 | -1.34 | .25 | 1.08 | .6 | 1.11 | .5 | .00 | 00154798 |
| 591 | 24 | 100 | -1.34 | .25 | 1.03 | .2 | 1.27 | 1.2 | .00 | 00146707 |
| 718 | 24 | 100 | -1.34 | .25 | .90 | -.8 | .79 | -1.1 | .00 | 00149040 |
| 1192 | 24 | 100 | -1.34 | .25 | 1.23 | 1.7 | **1.44** | **1.8** | .00 | 00158246 |
| 1216 | 24 | 100 | -1.34 | .25 | 1.15 | 1.2 | 1.16 | .7 | .00 | 00242313 |
| 1834 | 24 | 100 | -1.34 | .25 | .96 | -.3 | .99 | -.1 | .00 | 00166346 |
| 1914 | 24 | 100 | -1.34 | .25 | .81 | -1.6 | .83 | -.9 | .00 | 00143489 |
| 2181 | 24 | 100 | -1.34 | .25 | 1.14 | 1.1 | **1.42** | **1.8** | .00 | 00157901 |
| 2207 | 24 | 100 | -1.34 | .25 | 1.14 | 1.1 | **1.42** | **1.8** | .00 | 00157901 |
| 2225 | 24 | 100 | -1.34 | .25 | .98 | -.1 | .90 | -.5 | .00 | 00303205 |
| 205 | 23 | 100 | -1.40 | .25 | 1.03 | .2 | 1.26 | 1.1 | .00 | 00160717 |
| 631 | 23 | 100 | -1.40 | .25 | 1.03 | .2 | 1.21 | .9 | .00 | 00146214 |
| 895 | 23 | 100 | -1.40 | .25 | 1.22 | 1.6 | **1.50** | **2.0** | .00 | 00246279 |
| 1629 | 23 | 100 | -1.40 | .25 | 1.24 | 1.8 | **1.56** | **2.2** | .00 | 00165627 |
| 1676 | 23 | 100 | -1.40 | .25 | 1.18 | 1.3 | **1.38** | 1.6 | .00 | 00163705 |
| 1694 | 23 | 100 | -1.40 | .25 | 1.18 | 1.3 | 1.38 | 1.6 | .00 | 00163705 |
| 1954 | 23 | 100 | -1.40 | .25 | 1.14 | 1.1 | 1.19 | .8 | .00 | 00250153 |
| 2109 | 23 | 100 | -1.40 | .25 | 1.14 | 1.1 | 1.06 | .3 | .00 | 00305298 |
| 160 | 22 | 100 | -1.47 | .26 | **1.35** | **2.4** | **1.58** | **2.2** | .00 | 00159485 |
| 337 | 22 | 100 | -1.47 | .26 | 1.24 | 1.7 | **1.51** | **2.0** | .00 | 00149040 |
| 396 | 22 | 100 | -1.47 | .26 | 1.18 | 1.3 | 1.26 | 1.0 | .00 | 00248934 |
| 623 | 22 | 100 | -1.47 | .26 | 1.04 | .3 | 1.16 | .7 | .00 | 07701097 |
| 637 | 22 | 100 | -1.47 | .26 | 1.10 | .7 | 1.16 | .7 | .00 | 00152111 |
| 1253 | 22 | 100 | -1.47 | .26 | 1.06 | .4 | 1.26 | 1.1 | .00 | 00162349 |
| 1262 | 22 | 100 | -1.47 | .26 | 1.05 | .4 | 1.06 | .3 | .00 | 00246881 |
| 1705 | 22 | 100 | -1.47 | .26 | **1.30** | **2.1** | **1.44** | 1.7 | .00 | 06061614 |
| 1743 | 22 | 100 | -1.47 | .26 | .91 | -.7 | .86 | -.6 | .00 | 00252258 |
| 260 | 21 | 100 | -1.54 | .26 | 1.20 | 1.4 | 1.29 | 1.1 | .00 | 00246279 |
| 1309 | 21 | 100 | -1.54 | .26 | 1.14 | 1.0 | 1.21 | .8 | .00 | 00252671 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1486 | 21 | 100 | -1.54 | .26 | 1.07 | .5 | 1.08 | .3 | .00 | 00157230 |
| 2176 | 21 | 100 | -1.54 | .26 | 1.22 | 1.5 | **1.38** | 1.4 | .00 | 00152129 |
| 2202 | 21 | 100 | -1.54 | .26 | 1.22 | 1.5 | **1.38** | 1.4 | .00 | 00152129 |
| 5 | 20 | 100 | -1.61 | .27 | 1.12 | .8 | **1.35** | 1.3 | .00 | 00240061 |
| 893 | 20 | 100 | -1.61 | .27 | 1.20 | 1.3 | **1.61** | **2.1** | .00 | 00245411 |
| 1854 | 20 | 100 | -1.61 | .27 | 1.21 | 1.4 | **1.63** | **2.1** | .00 | 00240120 |
| 334 | 19 | 100 | -1.68 | .27 | .96 | -.3 | .83 | -.7 | .00 | 00145668 |
| 896 | 19 | 100 | -1.68 | .27 | .99 | -.1 | .91 | -.4 | .00 | 00246376 |
| 2 | 18 | 100 | -1.76 | .28 | 1.03 | .2 | 1.08 | .3 | .00 | 00163227 |
| 598 | 18 | 100 | -1.76 | .28 | 1.25 | 1.5 | **1.76** | **2.3** | .00 | 00152926 |
| 1187 | 18 | 100 | -1.76 | .28 | 1.26 | 1.5 | **1.65** | **2.0** | .00 | 00153932 |
| 634 | 17 | 100 | -1.83 | .28 | 1.09 | .5 | 1.23 | .8 | .00 | 00149300 |
| 1702 | 17 | 100 | -1.83 | .28 | 1.18 | 1.1 | 1.32 | 1.0 | .00 | 06061601 |
| 2186 | 17 | 100 | -1.83 | .28 | 1.09 | .5 | 1.03 | .1 | .00 | 00161364 |
| 2212 | 17 | 100 | -1.83 | .28 | 1.09 | .5 | 1.03 | .1 | .00 | 00161364 |
| 174 | 14 | 100 | -2.09 | .30 | .89 | -.6 | .72 | -1.0 | .00 | 00161369 |
| MEAN | 57. | 100. | .45 | .25 | 1.00 | -.1 | 1.00 | .0 | | |
| S.D. | 20. | 0. | 1.17 | .06 | .09 | 1.0 | .21 | 1.0 | | |

## ANEXO 6.2. VALIDACION DE CONSTRUCTO
### Análisis de la información de salida del patrón de respuestas al EXEDII.

Análisis Factorial Exploratorio.

En este anexo se incluye el reporte de salida del Análisis Factorial Exploratorio efectuado a partir de los resultados obtenidos en el EXEDII por las cohortes de 2005 y 2006.

El análisis no sigue una estrategia estándar, sino que se requiere tomar una serie de decisiones en cada paso a calcular dependiendo de los objetivos del análisis. Para efectuar el AFE se elaboró el siguiente archivo de instrucciones, adaptado del que utilizó por González Montesinos (2004) en su tesis doctoral:

```
>TITLE
TodoDefin.TSF - EFA DATA FULL-INFORMATION ITEM FACTOR ANALYSIS

>PROBLEM    NITEMS=100, RESPONSE=3;
>COMMENTS
                EXEDII Defin
                 Full-information item factor analysis
                 VARIMAX rotation
                 Data layout: COLUMN1 TO 100 Item Responses

        >NAMES
                ITEM1,ITEM2,ITEM3,ITEM4,ITEM5,ITEM6,ITEM7,ITEM8,
                ITEM9, ITEM10,ITEM11,ITEM12,ITEM13,ITEM14,ITEM15,
                ITEM16,ITEM17,ITEM18,ITEM19,ITEM20,ITEM21,ITEM22,
                ITEM23,ITEM24,ITEM25,ITEM26,ITEM27,ITEM28,ITEM29,
                ITEM30,ITEM31,ITEM32,ITEM33,ITEM34,ITEM35,ITEM36,
                ITEM37,ITEM38,ITEM39,ITEM40,ITEM41,ITEM42,ITEM43,
                ITEM44,ITEM45,ITEM46,ITEM47,ITEM48,ITEM49,ITEM50,
                ITEM51,ITEM52,ITEM53,ITEM54,ITEM55,ITEM56,ITEM57,
                ITEM58,ITEM59,ITEM60,ITEM61,ITEM62,ITEM63,ITEM64,
                ITEM65,ITEM66,ITEM67,ITEM68,ITEM69,ITEM70,ITEM71,
                ITEM72,ITEM73,ITEM74,ITEM75,ITEM76,ITEM77,ITEM78,
                ITEM79,ITEM80,ITEM81,ITEM82,ITEM83,ITEM84,ITEM85,
                ITEM86,ITEM87,ITEM88,ITEM89,ITEM90,ITEM91,ITEM92,
                ITEM93,ITEM94,ITEM95,ITEM96,ITEM97,ITEM98,ITEM99,ITEM100;
        >RESPONSE ' ','0','1';
```

```
>KEY 11111111111111111111111111111111111111111111111111111111111111111
   11111111111111111111111111111111111111111;

>RELIABILITY KR2;
>TETRACHORIC NDEC=3, LIST;

>FACTOR  NFAC=3, NROOT=6, ROTATE=VARIMAX, RESIDUAL, SMOOTH;
>FULL    CYCLES=20;
>SAVE    SMOOTH, ROTATED, PARM;
>INPUT   FILE='TodoD.dat' NIDCHAR=10;
(10A1,T1,100A1)
>STOP;
```

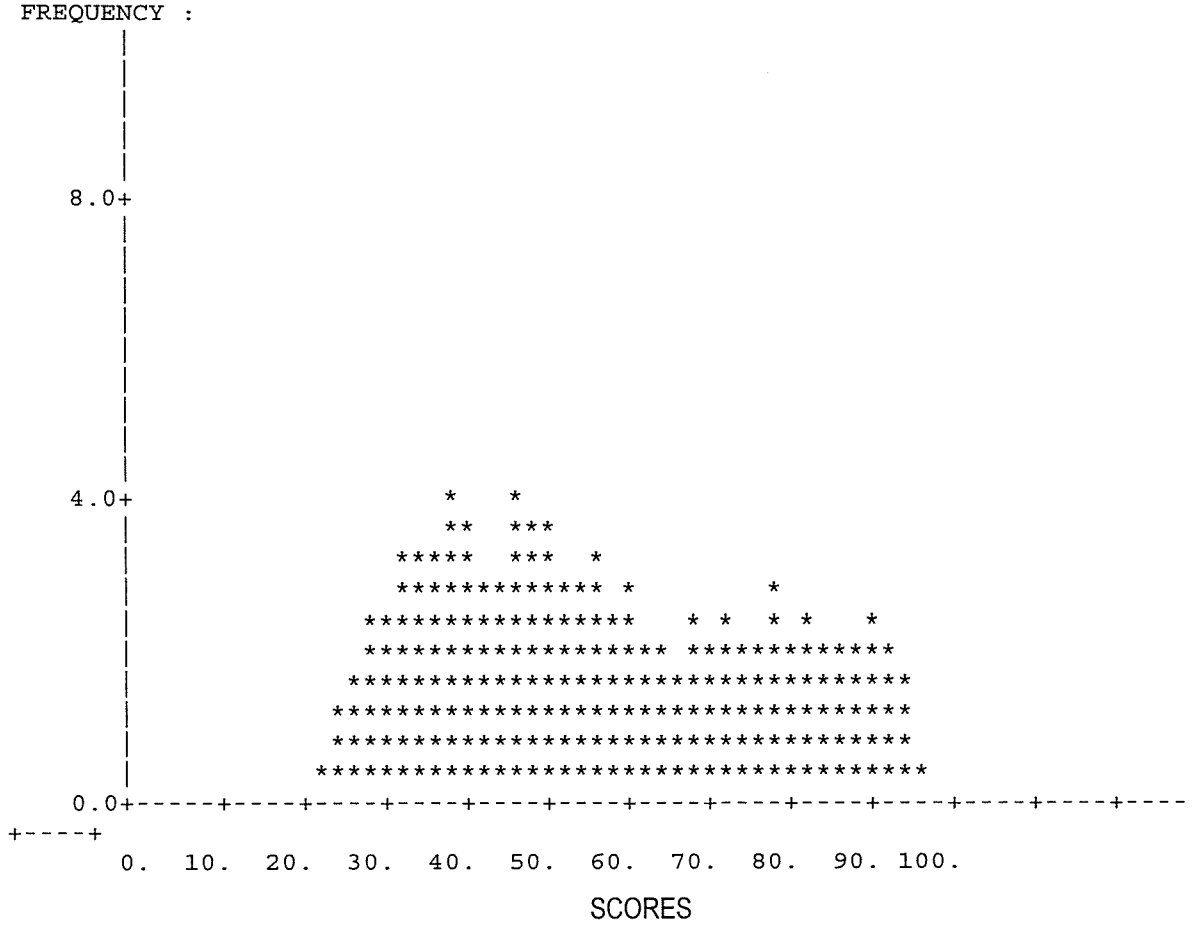
Enseguida se muestra el histograma que se proporciona en el cual se observa que, graficando en el eje de las X los puntajes, y en el eje de las Y las frecuencias de éstos se dibuja una curva semejante a la normal, con varios picos, correspondientes a un amplio rango de puntajes, desde 20 hasta 98.

PHASE 1: HISTOGRAM AND BASIC STATISTICS

TodoDefin.TSF - EFA DATA FULL-INFORMATION ITEM FACTOR ANALYSIS

---

MAIN TEST HISTOGRAM




Después de mostrar el número de observaciones por puntaje, o tabla de frecuencias, se muestran la media y la desviación estándar de los puntajes totales, que son:

| TEST | RECORD | NUMBER | MEAN | S.D. | PROPORTION | S.D. |
|---|---|---|---|---|---|---|
| MAIN | 2260 | 2260 | 57.0 | 20.3 | 0.570 | 0.495 |

En la Fase, la 2 se ofrecen las estadísticas básicas, que se muestran en una tabla que contiene nueve columnas: reactivo (número), número de respuestas, media, desviación estándar, RMEAN (que es el puntaje promedio de aquellos sustentantes que respondieron correctamente el reactivo), la facilidad del

reactivo (que es la proporción de respuestas correctas al reactivo), la dificultad del reactivo (o estadístico delta), y las dos últimas columnas son la correlación biserial y la punto biserial. Esta última es la correlación del puntaje al reactivo y el puntaje total. Se muestra un fragmento del archivo de la tabla:

```
PHASE 2: ITEM STATISTICS

 TodoDefin.TSF - EFA DATA FULL-INFORMATION ITEM FACTOR ANALYSIS

 -------------------------------------------------------------

     MAIN TEST ITEM STATISTICS
ITEM        NUMBER    MEAN    S.D.   RMEAN FACILITY    DIFF      BIS    P.BIS
1 ITEM1       2260   57.05   20.26   61.12    0.761   10.17    0.492    0.358
2 ITEM2       2260   57.05   20.26   65.68    0.582   12.17    0.635    0.503
3 ITEM3       2260   57.05   20.26   63.62    0.658   11.38    0.580    0.449
4 ITEM4       2260   57.05   20.26   66.48    0.389   14.12    0.473    0.372
5 ITEM5       2260   57.05   20.26   73.15    0.348   14.56    0.749    0.581
6 ITEM6       2260   57.05   20.26   67.13    0.533   12.67    0.668    0.532
7 ITEM7       2260   57.05   20.26   66.03    0.607   11.91    0.700    0.551
8 ITEM8       2260   57.05   20.26   65.74    0.481   13.19    0.518    0.413
9 ITEM9       2260   57.05   20.26   57.76    0.884    8.23    0.159    0.097
10 ITEM10     2260   57.05   20.26   59.64    0.882    8.25    0.572    0.350
11 ITEM11     2260   57.05   20.26   64.76    0.669   11.25    0.702    0.541
12 ITEM12     2260   57.05   20.26   64.19    0.346   14.58    0.330    0.256
```

Enseguida se muestra el estimado de consistencia interna o confiabilidad, el cual arroja un valor de 0.956, considerado alto.

```
KUDER-RICHARDSON KR20 ESTIMATE OF INTERNAL CONSISTENCY
   MAIN TEST KR20 = 0.956
```

En la Fase 5 se revisan las correlaciones tetracóricas. Primeramente se muestran columnas que presentan la información de respuestas omitidas. En el presente estudio, no hay respuestas faltantes. Se muestra un fragmento de la tabla del archivo de salida. La tabla que resume estos resultados tiene seis columnas: número de reactivo, de número de casos, porcentaje de respuestas correctas, porcentaje de respuestas omitidas, porcentaje de respuestas no alcanzadas y porcentaje de respuestas no contestadas.

```
PHASE 5:  TETRACHORIC CORRELATIONS
De
 TodoDefin.TSF - EFA DATA FULL-INFORMATION ITEM FACTOR ANALYSIS

 -------------------------------------------------------------

  MAIN TEST MISSING RESPONSE INFORMATION
```

| ITEM | NUMBER OF CASES | PERCENT CORRECT | PERCENT OMITTED | PERCENT NOT REACHED | PERCENT NOT PRESENTED |
|---|---|---|---|---|---|
| 1. ITEM1 | 2260 | 76.1 | 0.0 | 0.0 | 0.0 |
| 2. ITEM2 | 2260 | 58.2 | 0.0 | 0.0 | 0.0 |
| 3. ITEM3 | 2260 | 65.8 | 0.0 | 0.0 | 0.0 |
| 4. ITEM4 | 2260 | 38.9 | 0.0 | 0.0 | 0.0 |
| 5. ITEM5 | 2260 | 34.8 | 0.0 | 0.0 | 0.0 |
| 6. ITEM6 | 2260 | 53.3 | 0.0 | 0.0 | 0.0 |
| 7. ITEM7 | 2260 | 60.7 | 0.0 | 0.0 | 0.0 |
| 8. ITEM8 | 2260 | 48.1 | 0.0 | 0.0 | 0.0 |
| 9. ITEM9 | 2260 | 88.4 | 0.0 | 0.0 | 0.0 |
| 10.ITEM10 | 2260 | 88.2 | 0.0 | 0.0 | 0.0 |

A partir de este punto, el archivo de salida presenta la información en desplegados (Displays) numerados. Se utilizará el término en inglés por considerarse un término técnico.

Display No. 1 presenta la Matriz de correlaciones tetracóricas. La correlación tetracórica es ampliamente utilizada como medida de asociación entre dos reactivos dicotómicos. La correlación se obtiene sobre el supuesto de que existe una variable continua latente que subyace a cada reactivo, al que se le ha impuesto una calificación dicotómica. También se hace la suposición de que, para cada par de reactivos, las dos variables latentes correspondientes tienen una distribución normal, bivariada.

```
AVERAGE TETRACHORIC CORRELATION =  0.2941
     STANDARD DEVIATION =   0.1241
     NUMBER OF VALID ITEM PAIRS =   4950
```

DISPLAY 1. TETRACHORIC CORRELATION MATRIX

| | 1 ITEM1 | 2 ITEM2 | 3 ITEM3 | 4 ITEM4 | 5 ITEM5 | 6 ITEM6 | |
|---|---|---|---|---|---|---|---|
| 1ITEM1 | 1.000 | | | | | | |
| 2ITEM2 | 0.301 | 1.000 | | | | | |
| 3ITEM3 | 0.366 | 0.427 | 1.000 | | | | |
| 4ITEM4 | 0.230 | 0.333 | 0.283 | 1.000 | | | |
| 5ITEM5 | 0.310 | 0.519 | 0.492 | 0.336 | 1.000 | | |
| 6ITEM6 | 0.256 | 0.500 | 0.407 | 0.345 | 0.584 | 1.000 | |
| 7ITEM7 | 0.339 | 0.453 | 0.406 | 0.314 | 0.509 | 0.500 | |
| 8ITEM8 | 0.311 | 0.394 | 0.314 | 0.243 | 0.442 | 0.413 | |
| 9ITEM9 | 0.151 | 0.093 | 0.113 | 0.062 | 0.016 | 0.030 | |
| 10ITEM10 | 0.303 | 0.424 | 0.329 | 0.336 | 0.475 | 0.413 | |
| 11ITEM11 | 0.374 | 0.502 | 0.452 | 0.328 | 0.511 | 0.500 | |
| 12ITEM12 | 0.086 | 0.254 | 0.264 | 0.147 | 0.279 | 0.226 | |
| 13ITEM13 | 0.305 | 0.420 | 0.352 | 0.279 | 0.489 | 0.396 | |
| 14ITEM14 | 0.192 | 0.231 | 0.181 | 0.088 | 0.304 | 0.249 | |

A partir de la Fase 6 se presenta el Análisis Factorial Exploratorio. Primeramente se muestran los factores extraídos y las raíces latentes positivas de la matriz de correlación.

```
PHASE 6:  FACTOR ANALYSIS

 TodoDefin.TSF - EFA DATA FULL-INFORMATION ITEM FACTOR ANALYSIS

 ------------------------------------------------------------
    NUMBER OF FACTORS =            3
    FULL INFORMATION
    ADAPTIVE QUADRATURE

 DISPLAY    2.   THE POSITIVE LATENT ROOTS OF THE CORRELATION MATRIX

              1          2          3          4          5          6

     1     32.09832   3.89147    2.89154    2.56961    1.64417    1.46233


              7          8          9         10         11         12

     1      1.35532    1.32278    1.26497    1.22585    1.22001    1.17181


             13         14         15         16         17         18

     1      1.14550    1.13002    1.12181    1.07029    1.05631    1.02764
```

En el fragmento de tabla que se muestra arriba se puede notar que se obtuvieron cuatro raíces latentes positivas con información interpretable. Se marcan con rojo.

Por regla general se considera que las cargas de .30 o mayores indican alineación del reactivo con el factor, siempre que las cargas en los otros factores sean menores que ese valor. Mientras más alta sea la carga de un reactivo en un factor se considera que existe una mayor asociación entre el reactivo y el constructo que mide (González Montesinos, 2004).

```
DISPLAY    3NUMBER OF ITEMS AND SUM OF LATENT ROOTS
                  AND THEIR RATIO
                 100      100.0000000        1.0000000.
```

En el Display 4 se presenta la matriz de correlación suavizada, de la cual se muestra un fragmento a continuación:

DISPLAY 4. INITIAL SMOOTHED INTER-ITEM CORRELATION MATRIX

| | 1 ITEM1 | 2 ITEM2 | 3 ITEM3 | 4 ITEM4 | 5 ITEM5 | 6 ITEM6 |
|---|---|---|---|---|---|---|
| 1ITEM1 | 1.000 | | | | | |
| 2ITEM2 | 0.301 | 1.000 | | | | |
| 3ITEM3 | 0.366 | 0.427 | 1.000 | | | |
| 4ITEM4 | 0.230 | 0.333 | 0.283 | 1.000 | | |
| 5ITEM5 | 0.310 | 0.519 | 0.492 | 0.336 | 1.000 | |
| 6ITEM6 | 0.256 | 0.500 | 0.407 | 0.345 | 0.584 | 1.000 |
| 7ITEM7 | 0.339 | 0.453 | 0.406 | 0.314 | 0.509 | 0.500 |
| 8ITEM8 | 0.311 | 0.394 | 0.314 | 0.243 | 0.442 | 0.413 |
| 9ITEM9 | 0.151 | 0.093 | 0.113 | 0.062 | 0.016 | 0.030 |
| 10ITEM10 | 0.303 | 0.424 | 0.329 | 0.336 | 0.475 | 0.413 |
| 11ITEM11 | 0.374 | 0.502 | 0.452 | 0.328 | 0.511 | 0.500 |

En el Display 5 se presentan los estimados de las comunalidades iteradas. Las comunalidades se definen como el cuadrado de la correlación múltiple entre una variable observada y un grupo de factores. Es una forma de estimar la proporción de varianza en la variable observada, que es explicada por los factores extraídos que son comunes a la variable. Las cargas factoriales representan las correlaciones entre la variable observada y los factores extraídos. El método MINRES se utiliza a partir de este punto para obtener estimados más precisos.

DISPLAY 5. ITERATED COMMUNALITY ESTIMATES

| | | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 1 | ITEM1 | 0.271 | 0.246 | 0.245 | 0.245 |
| 2 | ITEM2 | 0.461 | 0.447 | 0.447 | 0.447 |
| 3 | ITEM3 | 0.368 | 0.349 | 0.348 | 0.348 |
| 4 | ITEM4 | 0.235 | 0.224 | 0.224 | 0.224 |
| 5 | ITEM5 | 0.594 | 0.594 | 0.596 | 0.596 |
| 6 | ITEM6 | 0.482 | 0.475 | 0.476 | 0.476 |
| 7 | ITEM7 | 0.495 | 0.484 | 0.484 | 0.484 |
| 8 | ITEM8 | 0.347 | 0.328 | 0.328 | 0.328 |
| 9 | ITEM9 | 0.052 | 0.034 | 0.033 | 0.033 |
| 10 | ITEM10 | 0.410 | 0.374 | 0.370 | 0.369 |
| 11 | ITEM11 | 0.544 | 0.526 | 0.524 | 0.523 |

En el Display 6 se presentan las raíces latentes más grandes de la matriz de correlación (NRoot). Se resaltan en rojo las tres más interpretables.

DISPLAY 6. THE NROOT LARGEST LATENT ROOTS OF THE CORRELATION MATRIX

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 31.54006 | 3.34873 | 2.33403 | 1.95966 | 0.93492 | 0.76679 |

En el Display 7 se presentan las principales cargas factoriales. TestFact utiliza los residuales mínimos al cuadrado (MINRES) para extraer los factores a partir de la matriz de correlación suavizada.

DISPLAY 7. MINRES PRINCIPAL FACTOR LOADINGS

| | | 1 | 2 | 3 |
|---|---|---|---|---|
| 1 | ITEM1 | 0.473 | -0.014 | 0.144 |
| 2 | ITEM2 | 0.621 | 0.148 | 0.197 |
| 3 | ITEM3 | 0.562 | 0.042 | 0.173 |
| 4 | ITEM4 | 0.465 | 0.058 | 0.066 |
| 5 | ITEM5 | 0.746 | 0.125 | 0.152 |
| 6 | ITEM6 | 0.652 | 0.182 | 0.133 |
| 7 | ITEM7 | 0.682 | 0.088 | 0.106 |
| 8 | ITEM8 | 0.506 | 0.158 | 0.218 |
| 9 | ITEM9 | 0.137 | -0.017 | 0.118 |
| 10 | ITEM10 | 0.558 | 0.128 | 0.203 |
| 11 | ITEM11 | 0.685 | 0.111 | 0.205 |

En el Display 8. Se presentan los estimados de Intercepto inicial y pendiente. El intercepto y la pendiente son funciones de la facilidad del reactivo y de las cargas factoriales. Si se omite la especificación de Rotación, entonces las cargas factoriales son las cargas factoriales MINRES que aparecen en este display. De lo contrario, se utilizan las cargas factoriales iniciales rotadas. Estos valores se utilizan como estimados iniciales para el procedimiento de máxima verosimilitud con información completa, que se especifica con el comando FULL.

DISPLAY 8. INITIAL INTERCEPT AND SLOPE ESTIMATES

| | | INTERCEPT | SLOPES 1 | 2 | 3 |
|---|---|---|---|---|---|
| 1 | ITEM1 | 0.815 | 0.461 | 0.286 | 0.173 |
| 2 | ITEM2 | 0.279 | 0.802 | 0.244 | 0.325 |
| 3 | ITEM3 | 0.502 | 0.620 | 0.303 | 0.241 |
| 4 | ITEM4 | -0.319 | 0.425 | 0.205 | 0.258 |
| 5 | ITEM5 | -0.614 | 1.010 | 0.414 | 0.534 |
| 6 | ITEM6 | 0.115 | 0.814 | 0.229 | 0.440 |
| 7 | ITEM7 | 0.378 | 0.773 | 0.365 | 0.456 |
| 8 | ITEM8 | -0.058 | 0.655 | 0.142 | 0.200 |
| 9 | ITEM9 | 1.213 | 0.162 | 0.085 | -0.020 |
| 10 | ITEM10 | 1.494 | 0.693 | 0.211 | 0.246 |
| 11 | ITEM11 | 0.633 | 0.911 | 0.355 | 0.376 |

En el Display 9 se presenta la estimación de los parámetros EM. Las estimaciones de los parámetros se basan en el método *Expectation Maximization* y en los puntos de cuadratura. Cuadratura es un método de integración numérica utilizado con frecuencia en la práctica para calcular el valor de una integral cuando no existe una solución de forma cerrada.

DISPLAY 9. THE EM ESTIMATION OF PARAMETERS

4 QUADRATURE POINTS

En el Display 10 se presentan los puntos de cuadratura y los pesos. Debe notarse que los pesos siempre son positivos y los puntos de cuadratura son simétricos:

DISPLAY 10. 4 QUADRATURE POINTS AND WEIGHTS:

| | | |
|---|---|---|
| 1 | -2.334414 | 0.045876 |
| 2 | -0.741964 | 0.454124 |
| 3 | 0.741964 | 0.454124 |
| 4 | 2.334414 | 0.045876 |

A partir de este punto, el archivo de salida muestra el desarrollo del procedimiento de iteración. En cada ciclo se reportan -2x Log likelihood y el máximo cambio en los valores del intercepto y la pendiente.

El procedimiento de iteración empieza con los valores iniciales de la pendiente y se van reportando las diferencias entre estos valores y los estimados revisados en cada ciclo.

Los pequeños pero máximos cambios en los estimados de la pendiente y del intercepto son una indicación de convergencia. Por ello, se van revisando los cambios en la pendiente hasta que se llegan a valores mínimos (de milésimas como mínimo en este tipo de análisis). Se muestra el primer ciclo y los dos últimos para ilustrar el proceso:

```
CYCLE  1  - 2 X MARGINAL LOG LIKELIHOOD =      0.2378780306D+06

              MAXIMUM CHANGE OF ESTIMATES
              INTERCEPT =    0.163317  SLOPE =    0.196963
                                                  0.176395
                                                  0.233712


  Number of patterns with zero probability =     0

 SUM OF MARGINAL PROBABILITIES =      0.29224D-04
.....

CYCLE 19  - 2 X MARGINAL LOG LIKELIHOOD =      0.2371690475D+06
                                    CHANGE =      0.2018340294D+01

              MAXIMUM CHANGE OF ESTIMATES
              INTERCEPT =    0.003273  SLOPE =    0.013835
                                                  0.001804
                                                  0.005304


  Number of patterns with zero probability =     0

 SUM OF MARGINAL PROBABILITIES =      0.35559D-04


    CYCLE 20  - 2 X MARGINAL LOG LIKELIHOOD =      0.2371652556D+06

              MAXIMUM CHANGE OF ESTIMATES
              INTERCEPT =    0.002974  SLOPE =    0.013420
                                                  0.001581
                                                  0.005165
```

En el Display 11 se presenta la $X^2$ y los grados de libertad. La Chi cuadrada es un estimador para juzgar si la inclusión de un parámetro adicional, o la remoción de un parámetro representa una mejoría significativa en el ajuste de los datos al modelo. Es interesante recordar que la Chi cuadrada es válida solamente cuando todos los patrones posibles de $2^n$ se han observado.

```
DISPLAY  11.      CHI-SQUARE =  202256.75  DF =     1862.00   P = 0.000

 CONVERGENCE NOT ATTAINED, CHI-SQUARE MAY BE INCORRECT
```

En este caso, el programa hace la indicación de que no se alcanzó la convergencia y que por ese motivo, la Chi cuadrada podría ser incorrecta. No obstante, conviene mencionar que una forma de cerciorarse que el programa trabajó bien y que los resultados obtenidos son confiables es la revisión de los datos que se han venido describiendo, además de verificar que, al final del archivo de salida aparezca la siguiente leyenda:

```
N O R M A L   E N D   O F   T H I S   P R O B L E M


  START DATE:  9-17-2007
  START TIME: 10:49:53
  END   TIME: 10:53:53
 NORMAL END
```

Así mismo, como se dijo anteriormente, se observó que el cambio máximo de los interceptos y pendientes llegó a las milésimas, por lo que se puede considerar que el valor de Chi cuadrada es correcto.

El siguiente paso del AFE consiste en realizar una prueba para investigar si el número de factores postulado en el primer análisis es el que más adecuadamente explica la estructura factorial del instrumento evaluado. Para ello fue necesario volver a "correr" el análisis con un número distinto de factores y verificar la mejor solución. Este proceso se presenta en el capítulo 6.

En el Display 12 se presentan los parámetros no transformados de los reactivos y los valores de los interceptos y las pendientes. A continuación se muestra un fragmento:

DISPLAY 12. UNTRANSFORMED ITEM PARAMETERS

| | | INTERCEPT | SLOPE ESTIMATES | | |
|---|---|---|---|---|---|
| | | | 1 | 2 | 3 |
| 1 | ITEM1 | 0.808 | 0.483 | 0.280 | 0.172 |
| 2 | ITEM2 | 0.288 | 0.829 | 0.260 | 0.211 |
| 3 | ITEM3 | 0.506 | 0.654 | 0.328 | 0.178 |
| 4 | ITEM4 | -0.331 | 0.396 | 0.163 | 0.184 |
| 5 | ITEM5 | -0.563 | 0.892 | 0.349 | 0.228 |
| 6 | ITEM6 | 0.130 | 0.858 | 0.264 | 0.240 |
| 7 | ITEM7 | 0.423 | 0.842 | 0.427 | 0.365 |
| 8 | ITEM8 | -0.076 | 0.659 | 0.140 | 0.056 |
| 9 | ITEM9 | 1.199 | 0.140 | 0.046 | -0.016 |
| 10 | ITEM10 | 1.702 | 0.979 | 0.322 | 0.368 |

En el Display 13 se presentan la dificultad y las comunalidades de los reactivos, así como los principales factores extraídos, como se muestra en un fragmento de la tabla:

DISPLAY 13. STANDARDIZED DIFFICULTY, COMMUNALITY, AND PRINCIPAL FACTORS

| | | DIFF. | COMM. | FACTORS 1 | 2 | 3 |
|---|---|---|---|---|---|---|
| 1 | ITEM1 | -0.698 | 0.254 | 0.492 | 0.057 | 0.094 |
| 2 | ITEM2 | -0.215 | 0.444 | 0.612 | 0.214 | 0.154 |
| 3 | ITEM3 | -0.404 | 0.362 | 0.575 | 0.101 | 0.146 |
| 4 | ITEM4 | 0.300 | 0.179 | 0.410 | 0.099 | 0.032 |
| 5 | ITEM5 | 0.401 | 0.492 | 0.657 | 0.180 | 0.170 |
| 6 | ITEM6 | -0.095 | 0.463 | 0.628 | 0.221 | 0.142 |
| 7 | ITEM7 | -0.297 | 0.506 | 0.697 | 0.117 | 0.082 |
| 8 | ITEM8 | 0.063 | 0.314 | 0.468 | 0.230 | 0.204 |
| 9 | ITEM9 | -1.186 | 0.022 | 0.115 | 0.044 | 0.080 |
| 10 | ITEM10 | -1.148 | 0.545 | 0.697 | 0.223 | 0.096 |
| 11 | ITEM11 | -0.458 | 0.554 | 0.710 | 0.179 | 0.135 |
| 12 | ITEM12 | 0.412 | 0.103 | 0.292 | 0.128 | 0.035 |

Los factores extraídos deben de explicar la mayor cantidad de varianza, por lo que se espera que no existan intercorrelaciones entre ellos. Para clarificar la relación de las variables con los factores en los que se agrupan, se realiza un procedimiento denominado Rotación, que puede ser de varios tipos. En el presente estudio se seleccionó la rotación Varimax que es la que maximiza la varianza observada entre las variables (Kachigan, 1991).

En el Display 14 se presenta el porcentaje de varianza de los tres factores extraídos. Es notable que el primer factor explica una mayor proporción de varianza que los otros dos.

DISPLAY 14. PERCENT OF VARIANCE

| | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 32.15754 | 3.51541 | 2.27874 |

En el Display 15 se presenta la matriz de correlación suavizada, positiva y definida.

DISPLAY 15. SMOOTHED CORRELATION MATRIX (POSITIVE-DEFINITE)

| | | 1 ITEM1 | 2 ITEM2 | 3 ITEM3 | 4 ITEM4 | 5 ITEM5 | 6 ITEM6 |
|---|---|---|---|---|---|---|---|
| 1 | ITEM1 | 1.000 | | | | | |
| 2 | ITEM2 | 0.328 | 1.000 | | | | |
| 3 | ITEM3 | 0.302 | 0.396 | 1.000 | | | |
| 4 | ITEM4 | 0.210 | 0.277 | 0.250 | 1.000 | | |
| 5 | ITEM5 | 0.349 | 0.467 | 0.420 | 0.292 | 1.000 | |
| 6 | ITEM6 | 0.335 | 0.454 | 0.404 | 0.284 | 0.476 | 1.000 |
| 7 | ITEM7 | 0.357 | 0.464 | 0.424 | 0.300 | 0.493 | 0.475 |
| 8 | ITEM8 | 0.263 | 0.367 | 0.322 | 0.221 | 0.384 | 0.374 |
| 9 | ITEM9 | 0.066 | 0.092 | 0.082 | 0.054 | 0.097 | 0.093 |
| 10 | ITEM10 | 0.365 | 0.489 | 0.437 | 0.311 | 0.514 | 0.501 |
| 11 | ITEM11 | 0.372 | 0.494 | 0.446 | 0.313 | 0.521 | 0.504 |

En el Display 16 se presenta una muestra de las correlaciones residuales.

DISPLAY 16. RESIDUAL CORRELATIONS

| | | 1 ITEM1 | 2 ITEM2 | 3 ITEM3 | 4 ITEM4 | 5 ITEM5 | 6 ITEM6 |
|---|---|---|---|---|---|---|---|
| 1 | ITEM1 | 0.746 | | | | | |
| 2 | ITEM2 | -0.027 | 0.556 | | | | |
| 3 | ITEM3 | 0.064 | 0.031 | 0.638 | | | |
| 4 | ITEM4 | 0.020 | 0.057 | 0.033 | 0.821 | | |
| 5 | ITEM5 | -0.039 | 0.052 | 0.071 | 0.044 | 0.508 | |
| 6 | ITEM6 | -0.079 | 0.046 | 0.003 | 0.061 | 0.108 | 0.537 |
| 7 | ITEM7 | -0.018 | -0.011 | -0.018 | 0.014 | 0.016 | 0.025 |
| 8 | ITEM8 | 0.049 | 0.027 | -0.008 | 0.022 | 0.059 | 0.039 |
| 9 | ITEM9 | 0.084 | 0.001 | 0.031 | 0.008 | -0.081 | -0.063 |
| 10 | ITEM10 | -0.062 | -0.065 | -0.109 | 0.025 | -0.039 | -0.087 |
| 11 | ITEM11 | 0.001 | 0.008 | 0.007 | 0.015 | -0.011 | -0.005 |

En el Display 17 se presentan los datos de la dificultad, las comunalidades y los factores extraídos con sus respectivas cargas factoriales, los cuales pueden ser interpretables después de la rotación. Esta tabla se presenta completa en el capítulo 6 de esta tesis.