

Validez de un prototipo de prueba de Español como Lengua Extranjera con Fines Académicos

Validity of a prototype test of Spanish as a Foreign Language for Academic Purposes

Ana Vine Jara

Universidad de Concepción
avine@udec.cl

Katia Sáez Carrillo

Universidad de Concepción
ksaez@udec.cl

RESUMEN

En el ámbito de la evaluación del Español como Lengua Extranjera con Fines Académicos (EFA) se requiere de estudios empíricos que den cuenta del proceso de validación de los instrumentos disponibles. En este artículo se presenta un estudio cuantitativo cuyo objetivo es evidenciar la validez de contenido y de apariencia de un prototipo de prueba EFA, en línea, para un nivel B2. Los resultados indican que se trata de un instrumento válido desde la perspectiva de un grupo de expertos y de una muestra de estudiantes extranjeros de intercambio universitario.

Palabras clave: validez, prueba en línea, Español como Lengua Extranjera, intercambio universitario

ABSTRACT

In the field of the assessment of Spanish as a Foreign Language for Academic Purposes (SFA), empirical studies are required to account for the validation process of the available instruments. In this article, a quantitative study is presented whose objective is to demonstrate the validity of content and appearance of an online SFA prototype test for a B2 level. The results indicate that this is a valid

instrument from the perspective of a group of experts and a sample of foreign exchange students.

Keywords: validity, online test, Spanish as a Foreign Language; university exchange

Fecha de recepción: 09/09/2017

Fecha de aprobación: 25/01/2018

1. INTRODUCCIÓN

Para que un instrumento de evaluación sea eficaz y proporcione la información requerida debe cumplir tres requisitos fundamentales: validez, confiabilidad y viabilidad (Alderson *et al.* 1998; MCER, 2002; Bordón, 2006). En la actualidad, se considera, además, que las pruebas o exámenes tienen que cumplir con criterios de autenticidad, de impacto y de ética y equidad (Bachman y Palmer, 1996; Llorián, 2012; Bordón, 2015). Sin embargo, la validez es uno de los temas que merece más atención en el contexto de la elaboración de exámenes.

En términos generales, la validez se refiere a si el examen, en su totalidad, mide lo que supone debe medir, lo cual se relaciona con los propósitos para los que fue creado el instrumento, por lo tanto, debe ser válido para esos objetivos. En consecuencia, se validan las inferencias relacionadas con el uso de un examen y no el propio instrumento, puesto que un test no es válido en sí mismo, sino para un propósito. De no cumplirse este requisito, entonces la prueba no es útil, ya que no se podrá hacer inferencias adecuadas de ella.

Para muchos usuarios de exámenes, la validez es una cualidad esencial en una prueba de lengua. Si un examen es válido, entonces, significa que es un buen examen. Desde la perspectiva de los usuarios, la validez de una prueba es una cualidad que otorgan los expertos en evaluación, acorde con los objetivos para los cuales fue diseñado el instrumento (Chapelle, 2017). De acuerdo con Martínez (2011), la validación es el proceso mediante el cual se determina si un

examen sirve y para qué sirve. La validez no es un concepto absoluto y definitivo, sino que es un argumento que requiere mayor o menor elaboración, dependiendo del uso que una determinada institución le otorgue a una prueba.

Mendoza (2015) realiza un interesante recorrido histórico del concepto de validez y de su aplicación práctica en el contexto educativo. Al respecto, señala que en la primera mitad del siglo XX, la validez se apoyaba en modelos estadísticos, los cuales se basaban en la correlación existente entre el puntaje obtenido por el candidato y el puntaje fijado como criterio. Generalmente, este criterio estaba vinculado con otros exámenes que evaluaban las mismas habilidades. Según Bachman (1990) y Mendoza (2015), Cronbach y Meehl (1955) introdujeron cuatro tipos de validez: predictiva, concurrente, de contenido y de constructo. Estos tipos se constituyeron en los pilares esenciales de la validez de exámenes hasta fines del siglo XX. El mayor aporte de Cronbach y Meehl fue concebir la validez como un proceso, denominado "validez de argumento", concepto que luego fue retomado por Kane (1992) y que consiste en recopilar evidencias (argumentos) que permitan realizar inferencias más que validar solo exámenes. En síntesis, la validación de un examen se considera como un proceso basado en la argumentación, esto debido a que los modelos anteriores no proporcionaban una guía adecuada que permitiera relacionar los resultados obtenidos en las pruebas y su interpretación de manera eficaz y equitativa.

La necesidad de diseñar pruebas de Español como Lengua Extranjera con Fines Académicos (EFA) se debe al creciente aumento de estudiantes extranjeros de intercambio universitario en distintas universidades de habla hispana (Pastor y Pandor, 2017). Precisamente, en Chile, un estudio del Ministerio de Educación (2016) demuestra que más de 7000 estudiantes extranjeros realizaron intercambios universitarios en el país durante el año 2015. De esta cifra, un alto número de alumnos proviene de países cuya lengua materna no es el español (Estados Unidos 26%, Francia 11% y Alemania 8%) y se prevé un aumento en el número de estudiantes de intercambio en los próximos años. Sin embargo, no existe una certificación obligatoria por parte de las universidades, que

acredite y exija un determinado nivel de español a los estudiantes extranjeros de intercambio universitario (Llorián, 2012; Ainciburu y Rodríguez, 2015). Esta preocupación y necesidad coincide con la de especialistas en universidades españolas y de otros países hispanohablantes (García *et al.* 2016; Mendoza, 2015; Quevedo, 2017; Pastor y Pandor, 2017).

El objetivo de este estudio es evidenciar la validez de contenido y de apariencia de un prototipo de prueba de EFA, en línea, de nivel B2. Para ello, en el segundo apartado del artículo se presentan los fundamentos teóricos que sustentan el trabajo; en el tercer apartado se describen los aspectos metodológicos; en el cuarto, los resultados y, finalmente, se dan a conocer las conclusiones del estudio.

2. LA VALIDEZ EN LOS INSTRUMENTOS DE EVALUACIÓN

Existen distintos tipos de validez, mientras Bordón (2006) se inclina por la validez interna porque no requiere de análisis estadísticos, Alderson *et al.* (1998) recomiendan que entre más tipos de validez se pueda establecer es mejor. Los diferentes tipos son en realidad distintos métodos para evaluar la validez (Bachman, 1990).

Alderson *et al.* (1998) clasifican la validez en interna, externa y de constructo. En la validez interna se considera la validez de apariencia, de contenido y de respuesta; y en la validez externa, la validez concurrente y predictiva, como se muestra en la figura 1. A continuación se definen los dos tipos de validación interna utilizados en el estudio: aparente y de contenido.

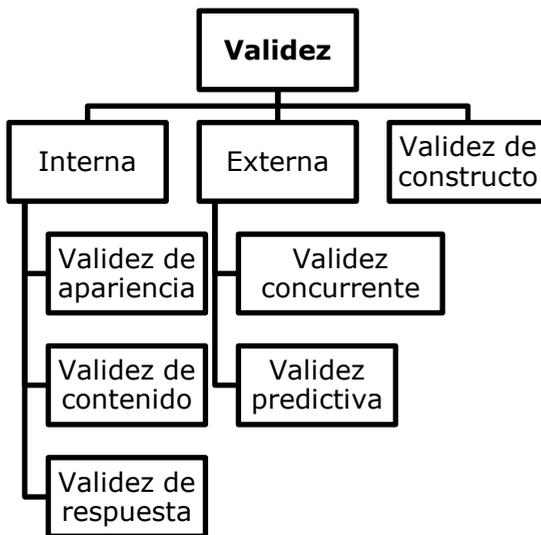


Figura 1. Tipos de validez (elaborada a partir de Alderson *et al.* 1998).

2.1 Validez de apariencia

La validez de apariencia se refiere a la opinión intuitiva de personas no expertas sobre una prueba. En el caso de la evaluación de segundas lenguas o lenguas extranjeras puede considerarse la opinión de hablantes nativos (no expertos en evaluación), así como también de los propios estudiantes de una segunda lengua a quienes se orienta el instrumento. Este tipo de validez se obtiene pidiendo a personas no expertas que rindan la prueba y luego den una opinión intuitiva respecto del instrumento, por ejemplo, una apreciación sobre las instrucciones, los ítems, la duración, la complejidad de las preguntas, el formato de aplicación, etc. Los resultados obtenidos permiten una primera apreciación en relación con la puesta en práctica de la prueba para hacer los ajustes necesarios antes de aplicarla al grupo meta. Este tipo de validación se considera poco científica por los expertos en evaluación. Sin embargo, en la actualidad, está adquiriendo relevancia, debido a la perspectiva comunicativa que prevalece en la evaluación de la lengua, lo cual supone que un examen comunicativo debería parecerse a lo que se hace en el mundo real con la lengua y, en este sentido, es

importante lo que piensan sobre el instrumento los usuarios de la lengua (Alderson *et al.* 1998; Martínez, 2011).

2.2 Validez de contenido

La validez de contenido se refiere a la representatividad o adecuación del contenido de un instrumento evaluativo, de acuerdo con los objetivos y el nivel al que se dirige. Se trata de comprobar si los ítems de una prueba están midiendo la dimensión que pretenden medir. Este tipo de validez requiere de la evaluación de expertos en el ámbito. Para ello se debe recoger opiniones de especialistas, por ejemplo, a través de escalas de valoración sobre los ítems de la prueba.

De acuerdo con Sireci (1998), la validez de contenido se puede estimar considerando dos perspectivas: el juicio de expertos y el uso de métodos estadísticos a partir de los datos obtenidos, luego de aplicar el instrumento a una muestra del grupo objeto.

El juicio de expertos se define como una opinión informada de personas con trayectoria en el tema objeto de estudio y que pueden dar información, evidencia, juicios y valoraciones (Escobar y Cuervo, 2008). En cuanto a la recopilación de la información señalada por los expertos existen métodos libres y otros más estructurados. En este sentido, es posible aplicar técnicas de juicio de expertos individuales o grupales. Según Escobar y Cuervo (2008), en la técnica individual, cada experto entrega su opinión por separado, mientras que en la técnica grupal se reúne a los jueces para lograr un nivel de acuerdo entre ellos. Respecto del número de expertos que se requiere, se observa una diversidad de opiniones entre los especialistas, Cabero y Llorente (2013) señalan que el número depende del acceso a ellos o la facilidad de contar con especialistas en el tema. Para realizar un juicio de expertos se debe elaborar cuestionarios o planillas donde el especialista pueda realizar la calificación. Al momento de la aplicación de este juicio, cada juez debe tener claro el objetivo de la evaluación y en qué consiste cada una de las dimensiones que va a evaluar.

2.3 La validez en las pruebas de EFA

En la actualidad se dispone de dos instrumentos que certifican el dominio en el EFA: el EXELEAA (Examen de Español como Lengua Extranjera para el Ámbito Académico), desarrollado por el Centro de Enseñanza para Extranjeros de la Universidad Nacional Autónoma de México y el eLADE (Examen en línea de Acreditación de Dominio de Español B1/B2), del Centro de Lenguas Modernas de la Universidad de Granada. Ambas pruebas tienen una data reciente, por lo que se desconoce el modelo de diseño, el análisis de necesidades lingüísticas, las etapas de validación, etc.

El EXELEAA es un examen multinivel, por lo que de acuerdo con el puntaje que obtiene el estudiante se le asigna el nivel de español que corresponda: B1, B2, C1. El objetivo de esta prueba es evaluar el nivel de competencia en español de los candidatos que desean cursar estudios superiores, tanto a nivel de pregrado como de posgrado. Al respecto, Mendoza (2015) menciona los lineamientos que han seguido para validar la tarea de escritura de la prueba, basándose en el enfoque de validación de exámenes de Kane (2006) y sustentado en la lógica argumentativa toulminiana. No obstante, se requiere considerar una serie de cuestiones relacionadas con una prueba para fines académicos, entre ellas, la representatividad de las tareas para el ámbito al cual se orientan, la diversificación de las prácticas de escritura y de tareas orales que varían de una disciplina a otra. Para dilucidar estos asuntos se necesita de estudios exploratorios que permitan conocer las principales necesidades de los estudiantes en contextos académicos y las diferencias más relevantes entre las distintas culturas académicas (Pastor, 2010; Llorián, 2012).

El eLADE es un examen binivel (B1/B2), en línea, que está dirigido a todos aquellos que desean obtener una acreditación de los niveles B1 o B2 de español por motivos profesionales o académicos. Martínez *et al.* (2015) describen acuciosamente el proceso de validación y los distintos análisis estadísticos que han llevado a cabo para validar el instrumento, utilizando análisis de IRT (Item Response Theory), con Winsteps. De esta manera, han logrado tener un

examen válido y que discrimine en los dos niveles que evalúa. No obstante, se debe tener en cuenta que esta prueba se orienta a dos fines: académicos y profesionales, y en Martínez *et al.* (2015) no se precisa cómo logran delimitar estos dos ámbitos, por ejemplo, no está claro si la diferenciación se realiza en los tipos de textos, en las tareas o los ítems o bien se cuenta con dos opciones de prueba, una para cada fin.

2.4 Ventajas de la implementación de pruebas de EFA a través del computador

La tecnología ha mostrado avances significativos en la enseñanza de lenguas extranjeras o segundas lenguas. Una muestra del impacto y los beneficios de la tecnología en el proceso de enseñanza y aprendizaje del Español como Lengua Extranjera se observa en los diferentes cursos que ofrece el Instituto Cervantes a través del Aula Virtual (Juan, 2009; Juan, 2015) o los cursos en línea del Programa de Español como Lengua Extranjera de la Universidad de Concepción (Ferreira *et al.* 2015). Esto, sin duda, es solo una muestra, pues en la actualidad un gran número de universidades y centros de idiomas cuenta con cursos implementados en entornos virtuales.

Pese al auge de la tecnología en la enseñanza del español como lengua extranjera, en la evaluación del español se han realizado pocos intentos de automatización de pruebas de dominio. De hecho, los exámenes de mayor reconocimiento internacional, como el DELE (Diploma de Español como Lengua Extranjera) y el CELU (Certificado de Español: Lengua y Uso) continúan aplicándose en formato de lápiz y papel. Sin embargo, el año 2016, el Instituto Cervantes, la Universidad Nacional Autónoma de México, la Universidad de Salamanca y la Universidad de Buenos Aires han puesto en marcha el examen de dominio SIELE (Servicio Internacional de Evaluación de la Lengua Española). Se trata de una prueba internacional de multinivel, en línea, que certifica el nivel de competencia en español mediante una escala de puntos equivalentes desde los niveles A1 al C1 según el MCER (2002). Su implementación en línea hace que el examen sea ágil y rápido en cuanto a la entrega de resultados; adaptable,

porque el usuario puede elegir si desea certificar su competencia global en español o en algunas de las habilidades lingüísticas; y homogéneo, puesto que evalúa las distintas variantes del español. Del mismo modo, los exámenes EXELEAA y eLADE, de data reciente, se encuentran implementados en un formato en línea. Es de esperar que la evaluación de Español como Lengua Extranjera continúe avanzando hacia formatos actualizados de aplicación de pruebas, acorde con las nuevas realidades educativas (García y Navarro, 2008).

La aplicación de pruebas en lápiz y papel tiene un costo elevado tanto para los estudiantes como para las instituciones. Este costo se observa en aspectos como el tiempo de entrega de resultados, la cantidad de correctores de pruebas, los examinadores para la evaluación oral, etc. Una prueba en línea representa una ventaja porque es menos costosa (Chapelle y Douglas, 2006). Además, mediante este formato de implementación se tiene flexibilidad en cuanto a los diferentes tipos de ítems, se puede evaluar globalmente las destrezas comunicativas, dar una corrección rápida y eficaz, y las pruebas orales se pueden realizar sin la presencia de los evaluadores, etc. Sin embargo, es necesario recordar que se debe utilizar la tecnología con responsabilidad y profesionalismo, sin olvidar los requisitos de validez, confiabilidad, viabilidad y equidad que debe cumplir todo instrumento de evaluación (Martínez, 2011).

2.5 Etapas en la evaluación de lenguas a través del computador

La evaluación de lenguas Asistida por computador (CALT, del inglés, *Computer Assisted Language Testing*) considera el uso de la tecnología para facilitar, contextualizar y mejorar la evaluación de las habilidades lingüísticas (Winke e Isbell, 2017).

Según Chapelle y Douglas (2006), la metodología de CALT, para la elaboración de pruebas válidas mediatizadas por computador, contempla dos etapas: 1) la implementación y 2) la evaluación de la prueba a partir de la aplicación.

2.5.1 La implementación en CALT

La implementación hace referencia a los recursos necesarios para elaborar una prueba, lo que dependerá de su propósito. Por ejemplo, las pruebas de alto impacto requieren de un conjunto de especificaciones y procedimientos para la selección de tareas adecuadas acorde con estas exigencias, así como también de métodos para recopilar, almacenar y analizar las respuestas de los candidatos. En principio se trabajaba con sistemas sofisticados que hacían del uso del computador algo costoso y poco alcanzable; pero en la actualidad se abordan propuestas sencillas, que permitan mayor accesibilidad para trabajar con esta metodología (Chapelle y Douglas, 2006). En términos generales, para implementar una prueba en línea se debe tener en cuenta:

- La creación de la prueba en la herramienta seleccionada.
- La publicación o visualización de la prueba en la web
- El acceso a la prueba por parte de los estudiantes
- La aplicación de la prueba
- La visualización de los resultados

2.5.2 La evaluación en CALT

Una vez que la prueba ha sido implementada bajo las directrices de la metodología CALT corresponde evaluarla. La evaluación en este contexto considera dos momentos: 1) la evaluación de la prueba en su modalidad computacional por parte de expertos en CALT y 2) la evaluación mediante la respuesta de quienes rinden la prueba en línea.

La evaluación de la prueba en su modalidad computacional, por parte de expertos en el ámbito, considera la navegación, el diseño de página, el texto y las barras de herramientas (Fulcher, 2003). Esto significa que la velocidad de navegación debe ser adecuada, los botones e íconos deben ser fáciles de usar y se debe garantizar las medidas de seguridad que corresponda. En cuanto al diseño de la página se debe presentar textos breves, reducir el desplazamiento (*scrolling*) en la pantalla, evitar el uso de letras mayúsculas y de animaciones que distraen. Además, se debe mostrar el menor número de barras de herramientas y posicionar en primer lugar la información relevante.

La evaluación por parte de los estudiantes permite detectar problemas relacionados con el funcionamiento de las instrucciones, la interfaz en los sitios y plataformas, así como también con la estructura de diseño computacional del instrumento, con el uso de botones de navegación, etc.

La preocupación actual en las pruebas asistidas por computador está relacionada con los conceptos de validez, confiabilidad e impacto (Ramak y Rahmattalab, 2017). En este sentido, la investigación en el ámbito requiere de programas de validación que permitan justificar las interpretaciones y usos de los resultados de las pruebas (Chapelle y Voss, 2016). En efecto, las pruebas computarizadas no solo deben aumentar la eficiencia en la administración y la puntuación de los exámenes, sino reflejar con precisión las formas en que los estudiantes han aprendido una segunda lengua (Chapelle y Douglas, 2006).

3. METODOLOGÍA

3.1 Tipo de estudio

En este estudio cuantitativo se presentan los resultados obtenidos en la validación de contenido y de apariencia del prototipo de prueba EFA, en línea, de nivel B2. La validez de contenido se llevó a cabo mediante la técnica de juicio de expertos en relación con 1) la dimensión que evalúan los ítems y 2) la implementación del instrumento en su modalidad computacional. La validez de apariencia se aplicó a una muestra de estudiantes de español como lengua extranjera de intercambio en una universidad chilena.

3.2 Objetivos del estudio

Los objetivos que orientan este estudio son:

1. Evidenciar la validez de contenido del prototipo de prueba EFA, en línea, de nivel B2 a través de un juicio de expertos.

2. Determinar la validez de apariencia del prototipo de prueba EFA, en línea, de nivel B2.

3.3 Descripción del prototipo de prueba EFA

La prueba de EFA, en línea, de nivel B2 se elaboró según el modelo de criterios evaluativos propuesto por Vine (2015) y Vine y Ferreira (2016). El modelo considera las directrices de las sociedades internacionales de evaluación de lenguas y de los especialistas en evaluación de segundas lenguas. Dicho modelo propone cuatro criterios: 1) éticos, 2) diseño, 3) elaboración, 4) validación. La prueba se orienta a un nivel B2, puesto que es el nivel esperable en el contexto de los intercambios académicos (Pastor, 2010).

Como se muestra en la tabla 1, la prueba consta de cinco secciones: comprensión lectora, comprensión auditiva, competencia lingüística, expresión escrita y expresión oral. Se debe precisar que los textos y audios que componen este prototipo de prueba fueron validados por un grupo de expertos.

Sección	Nº textos	Tipo y nº de ítems	Tiempo
Comprensión lectora	3 textos	V o F (1 texto con 5 preguntas)	30 minutos
		Selección múltiple con 4 opciones (2 textos con 5 preguntas cada uno)	
Comprensión auditiva	3 audios	V o F (1 audio con 5 preguntas)	30 minutos
		Selección múltiple con 4 opciones (2 audios con 5 preguntas cada uno)	
Competencia lingüística	1 texto	20 ítems de selección múltiple con 4 opciones	10 minutos
Expresión e interacción escrita	---	Escritura de un ensayo (se presentan dos opciones de temas, el estudiante elige uno)	20 minutos

Expresión e interacción oral	---	Entrevista: Consta de 3 partes: -Presentación del estudiante -Presentación oral sobre un tema a elección -Conversación sobre el tema	15 minutos
Duración total de la prueba			1 hora, 45 minutos

Tabla 1. Secciones del prototipo de prueba EFA

La sección de comprensión lectora se compone de tres textos y la sección auditiva de tres audios. En el primer texto de cada una de estas secciones se plantean cinco preguntas de verdadero y falso, y en los textos 2 y 3 se plantean cinco preguntas de selección múltiple por cada uno. En total, cada sección cuenta con 15 ítems de corrección objetiva. La sección de competencia lingüística comprende un texto con veinte ítems de selección múltiple donde se evalúan los diferentes usos gramaticales definidos para el nivel B2 en el contexto del EFA. En la expresión escrita se proponen dos temas: uno del área humanista y otro del área científica, el estudiante selecciona uno y escribe un ensayo. La expresión oral se divide en tres partes: 1) presentación del estudiante, 2) presentación de un tema a elección (selecciona tema del área humanista o científica) y argumentación sobre un punto de vista respecto de dicho tema y, finalmente, 3) conversación con el entrevistador sobre el tópico expuesto.

Para definir las dimensiones lingüísticas que se debe evaluar en una prueba de dominio en EFA de nivel B2 es pertinente considerar los descriptores del MCER (2002), según el nivel de lengua y las directrices de los especialistas en evaluación de segundas lenguas (Alderson *et al.* 1998; Martínez, 2011; Figueras y Puig, 2013).

Para medir las habilidades de comprensión se debe tener en cuenta tres estrategias básicas que se ponen en práctica en la comprensión de un texto y en las que un estudiante extranjero de intercambio universitario, de nivel

B2, debería mostrar dominio, se trata de: 1) la capacidad para seleccionar las palabras e ideas relevantes para comprender el texto; 2) la predicción o la capacidad que posee el lector para anticiparse al contenido de un texto, por ejemplo, el final de una historia, la lógica de una explicación, la estructura de una oración compleja, etc.; y 3) la inferencia o la capacidad para sacar deducciones y conclusiones acerca de lo que no está explícito en el texto (Jouini, 2005). Encontrar o descubrir el significado literal es la forma más elemental de la comprensión, mientras que las inferencias constituyen la esencia misma de este proceso.

En cuanto a la habilidad de expresión escrita en un contexto académico de nivel B2, el MCER (2002) señala que los alumnos deben ser capaces de sintetizar información, escribir textos donde desarrollen argumentos a favor o en contra de un punto de vista y explicar ventajas o desventajas sobre distintas opciones o planteamientos.

En la expresión oral, el alumno debe evidenciar la capacidad para realizar presentaciones orales sobre asuntos relacionados con su área de especialidad, desarrollando argumentos y defendiendo su punto de vista.

En síntesis, en este prototipo de prueba se evalúan seis dimensiones lingüísticas: 1) la identificación de información específica en textos especializados, 2) la inferencia de las ideas principales de los párrafos, 3) la inferencia de la idea global del texto, 4) los usos gramaticales correspondientes a un nivel B2, 5) el género ensayo y 6) el género presentación oral (MCER, 2002; Martínez, 2011; Figueras y Puig, 2013; Regueiro y Sáez, 2013).

Las secciones de comprensión lectora, auditiva, competencia lingüística y expresión escrita del prototipo de prueba se implementaron en la plataforma Moodle (*Object-Oriented Dynamic Learning Environment*) y la sección de expresión oral se aplicó de manera presencial.

3.4 Participantes

En la validación de contenido de la prueba participaron dos grupos de expertos, uno para evaluar los ítems del prototipo de prueba y el otro para valorar la implementación del

instrumento en línea. En la validación de los ítems del prototipo de prueba participaron ocho expertos, profesionales de lenguas (2 estudiantes de Magíster en Lingüística Aplicada, 1 Magíster en Lingüística, 3 candidatos al grado de Doctor en Lingüística y 2 Doctores en Lingüística). En la validación del instrumento en línea participaron 10 especialistas en la enseñanza y evaluación de lenguas con tecnología (3 estudiantes de Magíster en Lingüística Aplicada, 2 Magíster en Lingüística Aplicada, 3 candidatos al grado de Doctor en Lingüística y 2 Doctores en Lingüística).

En la selección de los expertos para evaluar los ítems del prototipo de prueba se consideró como requisito que los profesionales tuvieran un mínimo de 3 años de experiencia en el diseño de ítems para pruebas de lengua, con especialidad, en español para extranjeros. Del mismo modo, en la validación de la prueba en línea se tuvo en cuenta que cada uno de los profesionales tuviera una experiencia mínima de 3 años en enseñar y evaluar español como segunda lengua u otra lengua extranjera (principalmente, inglés) a través de plataformas educativas que permiten el diseño de cursos y pruebas.

En la validez de apariencia de la prueba participaron 28 estudiantes de español como lengua extranjera de intercambio universitario, cuyas lenguas maternas corresponden al francés 9 estudiantes (32%), al alemán 8 (28%), al inglés 4 (14%), al sueco 2 (7%), al portugués 2 (7%), al italiano 1 (4%), al checo 1 (4%) y al ruso 1 (4%).

3.5 Instrumentos

A continuación se describen los distintos cuestionarios que se elaboraron para evaluar la validez de contenido y de apariencia de la prueba de EFA, en línea, de nivel B2.

3.5.1 Cuestionario validez de contenido

Para recopilar los datos de la validación de contenido se elaboraron dos cuestionarios, uno sobre las dimensiones evaluadas por los ítems del prototipo de prueba (ver ejemplo en tabla 2 y ver cuestionario abreviado en anexo 1) y el otro sobre la modalidad de aplicación en línea (ver anexo 2). Se trata de cuestionarios de escalas Likert de 5 puntos (1: Muy en

desacuerdo, 2: En desacuerdo, 3: Indiferente, 4: De acuerdo, 5: Muy de acuerdo).

DIMENSIÓN III: INFERENCIA DE LA IDEA CENTRAL DEL TEXTO	Observación	INDICADORES				
		1	2	3	4	5
Preguntas sección comprensión de lectura						
<i>Texto I</i>						
5. A partir de la lectura del texto se puede decir que la sociedad requiere de egresados vinculados al ámbito laboral.						
<i>Texto II</i>						
5. ¿Cuál es la idea central del texto?						

Tabla 2. Ejemplo de cuestionario juicio de expertos sobre los ítems de la prueba

El cuestionario mediante el cual se evaluaron los ítems del prototipo de prueba considera seis dimensiones: identificación de información literal, inferencia de la idea principal a nivel de párrafo, inferencia de la idea global, competencia gramatical de nivel B2, género ensayo, género presentación oral. Estas dimensiones se obtuvieron de las propuestas del MCER (2002), Martínez (2011), Figueras y Puig (2013) y Regueiro y Sáez (2013).

El cuestionario sobre la modalidad de aplicación en línea se elaboró a partir de las directrices metodológicas de CALT (Chapelle y Douglas, 2006; Noijons, 1994). El objetivo de esta evaluación fue evidenciar en qué medida el prototipo de prueba en línea cumple con los principios metodológicos de CALT en lo referido a la implementación de una prueba de lengua. Dicho instrumento consta de dos dimensiones: enfoque evaluativo (5 ítems) y aspectos técnicos (10 ítems).

3.5.2 Cuestionario validez de apariencia

En la validación de apariencia también se diseñó un cuestionario de escala Likert de 5 puntos, con 12 ítems que evalúan la percepción de los estudiantes en relación con el

entorno de la prueba, la claridad de las instrucciones, la comprensión de los textos y audios y la duración del instrumento (ver figura 2 y anexo 3). Los ítems elaborados para este tipo de validación se basan en los lineamientos de Martínez (2011), en lo concerniente a las actitudes, reacciones e impresiones de los estudiantes hacia el prototipo de prueba.

3.6 Procedimiento

Luego de haber elaborado el prototipo de prueba EFA, en línea, de nivel B2 se aplicó el instrumento a un grupo de 9 estudiantes de español como lengua materna (estudiantes de la carrera de Traducción de una universidad chilena) para observar el funcionamiento de los ítems. Enseguida, y luego de corregir la versión del prototipo de prueba en línea (a partir de la información obtenida en la aplicación a estos 9 alumnos) se consideró la validez de contenido por un grupo de expertos y, finalmente, la validez de apariencia, a través de la percepción de una muestra de estudiantes extranjeros de intercambio universitario sobre el instrumento.

A partir de la aplicación a los estudiantes, hablantes nativos de español, se modificaron 12 ítems (8 de comprensión lectora y 4 de comprensión auditiva) que obtuvieron un alto porcentaje de aciertos (sobre el 70% de aciertos). Para ello, se elevó el nivel de dificultad de los distractores. Las instrucciones funcionaron adecuadamente, excepto, en la escritura, donde la forma en que se había planteado el tema no elicita las desventajas y, por consiguiente, la argumentación, sino que evidenciaba una visión positiva del tópico. Se realizaron estos cambios para así someter el instrumento a la evaluación de los expertos.

La validez de contenido de la prueba se llevó a cabo mediante la técnica de juicio de expertos. Un grupo de especialistas evaluó la prueba en relación con 1) la dimensión que evalúan los ítems acorde con el nivel B2 de EFA, y el otro grupo la evaluó en relación con 2) la implementación del instrumento en su modalidad computacional.

Para la validación de los ítems del prototipo de prueba, cada experto recibió un cuestionario tipo Likert donde se explicaba en qué consistía la evaluación, se proporcionaba la

lista de dimensiones con los correspondientes ítems y un usuario y una clave para acceder a la prueba en línea. Enseguida, leían/escuchaban los textos, leían las preguntas y evaluaban si los ítems medían las distintas dimensiones señaladas, y así sucesivamente con cada sección de la prueba, como se ejemplificó en la tabla 2.

En la validación del instrumento en su modalidad de aplicación en línea, se envió a cada experto un cuestionario tipo Likert y el usuario y clave para que accediera a la prueba. El cuestionario se dividió en dos dimensiones, la primera referida al enfoque evaluativo bajo el cual se elaboró la prueba, y la segunda a los aspectos técnicos de lo que implica plantear una prueba de lengua en línea, siguiendo las directrices metodológicas de CALT (ver anexo 2).

Con los datos de la validación de contenido se procedió a realizar algunas modificaciones indicadas por los expertos para, luego, aplicar el instrumento a una muestra de estudiantes extranjeros de intercambio en una universidad chilena. Después de que los estudiantes respondieron la prueba, se les aplicó el cuestionario de apreciación sobre el instrumento en su conjunto, con el objetivo de conocer su opinión, como usuarios de la lengua, en relación con la comprensibilidad de las instrucciones, de los textos, duración de la prueba, implementación en línea, etc.

4. RESULTADOS

Una vez obtenida la evaluación por parte de los expertos, se revisó el nivel de acuerdo de los jueces con la finalidad de determinar si los ítems medían las dimensiones que debían evaluar y si la implementación en línea del instrumento reflejaba los presupuestos teóricos y metodológicos establecidos por CALT. Enseguida, se aplicó el instrumento a una muestra de 28 estudiantes de español como lengua extranjera. A continuación se presentan los resultados de los dos tipos de validez considerados en el estudio: validez de contenido y de apariencia.

4.1 Validez de contenido de los ítems de la prueba

En la tabla 3 se presenta la evaluación de los expertos sobre los ítems de la prueba. En términos generales, se visualiza un alto nivel de acuerdo entre los jueces en los ítems de las distintas dimensiones de la prueba (promedio 81,1%, *muy de acuerdo*).

Dimensión	Texto	N	De acuerdo		Muy de acuerdo	
			n	%	n	%
DIMENSIÓN I	Texto I	1	0	0,0	8	100,0
		2	6	75,0	2	25,0
	Texto II	1	0	0,0	8	100,0
		2	0	0,0	8	100,0
	Texto III	1	0	0,0	8	100,0
		2	1	12,5	7	87,5
	Audio I	1	0	0,0	8	100,0
		2	2	25,0	6	75,0
	Audio II	1	2	25,0	6	75,0
		2	0	0,0	8	100,0
	Audio III	1	1	12,5	7	87,5
		2	2	25,0	6	75,0
DIMENSIÓN II	Texto I	3	1	12,5	7	87,5
		4	3	37,5	5	62,5
	Texto II	3	3	37,5	5	62,5
		4	2	25,0	6	75,0
	Texto III	3	2	25,0	6	75,0
		4	0	0,0	8	100,0
	Audio I	3	1	12,5	7	87,5
		4	2	25,0	6	75,0
	Audio II	3	0	0,0	8	100,0
		4	2	25,0	6	75,0
	Audio III	3	2	25,0	6	75,0
		4	2	25,0	6	75,0
DIMENSIÓN III	Texto I	5	1	12,5	7	87,5
	Texto II	5	3	37,5	5	62,5
	Texto III	5	3	37,5	5	62,5
	Audio I	5	0	0,0	8	100,0
	Audio II	5	1	12,5	7	87,5
	Audio III	5	2	25,0	6	75,0
DIMENSIÓN IV			2	25,0	6	75,0
DIMENSIÓN V			1	12,5	7	87,5
DIMENSIÓN VI			3	37,5	5	62,5
Promedio			18,9		81,1	

Tabla 3. Resultados validación de los ítems

Se observa en la tabla 3, que en la dimensión I de la comprensión lectora y auditiva, que corresponde a las preguntas 1 y 2 de cada texto, todos los jueces calificaron los ítems entre 4 y 5 en la escala Likert, lo que significa que hay acuerdo entre el grupo de expertos en señalar que los ítems están midiendo la dimensión de identificar información específica en los textos orales y escritos. De hecho, los 8 expertos coinciden en calificar 6 preguntas de esta dimensión con la categoría *muy de acuerdo*, lo que otorga seguridad en cuanto a que el instrumento está evaluando lo que corresponde.

En la dimensión II sobre la inferencia de las ideas principales de los párrafos, que pertenece a las preguntas 3 y 4 de comprensión lectora y auditiva, los ítems también fueron calificados entre 4 y 5 en la escala Likert. No obstante, un evaluador precisó que la forma en que estaban redactadas las preguntas 3 y 4 del texto 2 no elicitan la inferencia, aunque evaluaban la dimensión señalada. Se realizaron las modificaciones y el experto señaló estar de acuerdo con el cambio realizado. Los enunciados de estos ítems quedaron de la siguiente manera:

De acuerdo con el texto, la mejor forma de aprender la habilidad auditiva es... (pregunta 3, texto 2)

Según el texto, la habilidad motora de una segunda lengua... (pregunta 4, texto 2).

En cuanto a la dimensión III se observa que las preguntas planteadas evalúan la inferencia de la idea central de cada texto oral y escrito, puesto que todos los expertos situaron sus respuestas en las categorías *de acuerdo* y *muy de acuerdo*.

La dimensión IV, que corresponde al dominio de la competencia lingüística para el nivel B2 de EFA, presenta un alto nivel *de acuerdo* entre los evaluadores, 6 *muy de acuerdo* y 2 *de acuerdo* (75% y 25% respectivamente). No obstante, los resultados obtenidos en la aplicación del prototipo de prueba a la muestra de estudiantes extranjeros (ver figura 1)

evidencian la necesidad de modificar el tipo de ítem mediante el cual se evalúa la competencia gramatical.

En la dimensión V, 7 de los expertos señalan estar *muy de acuerdo* con la tarea que evalúa el género académico ensayo, acorde con el nivel B2. Dos de los expertos sugirieron que se podría abordar otros temas más académicos en la opción 2 de escritura que, en principio, trataba sobre "El día mundial sin compras". Se consideró la indicación de los expertos y se realizó la modificación con la finalidad de mejorar la calidad y validez del instrumento. La opción 2 de la tarea de escritura fue reemplazada por el tema de "La crisis del agua a nivel mundial".

En la dimensión VI se evalúa la capacidad del estudiante para realizar una presentación oral. En este sentido, los jueces evidencian altos niveles *de acuerdo* y *muy de acuerdo*, lo que indica que el ítem mide lo que dice medir.

En la figura 1 se muestra el desempeño obtenido por la muestra de estudiantes extranjeros de intercambio universitario en el prototipo de prueba. Se presentan estos resultados con el objetivo de contrastarlos con la evaluación de los expertos.

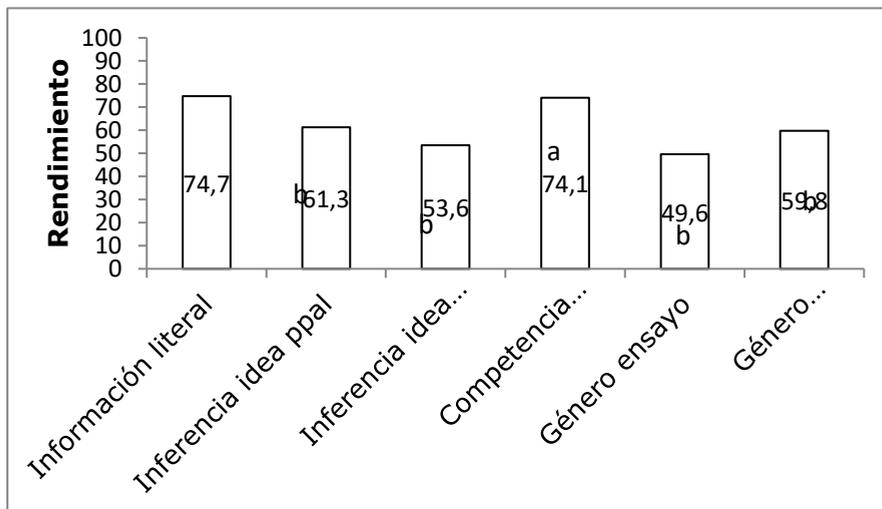


Figura 1. Resultados de la muestra en las distintas dimensiones de la prueba

Se observa en la figura 1 que las preguntas están midiendo lo que dicen medir, excepto, los ítems de competencia gramatical, que si bien evalúan el nivel B2 de EFA, como señalan los expertos, el tipo de ítem no es apropiado para elicitación de la competencia gramatical.

El rendimiento logrado en las dimensiones D1 (información literal) y D4 (competencia gramatical) fueron significativamente mayores a las restantes dimensiones ($F(5, 135)=11,5$; $p<0,0001$). Esto evidencia que ambas dimensiones resultaron fáciles para los estudiantes. Por un lado, la identificación de información literal en los textos/audios y, por otro, la competencia gramatical. El elevado número de aciertos en la competencia gramatical podría indicar, en principio, que los estudiantes tienen un buen dominio lingüístico; sin embargo, en las tareas de expresión escrita y oral mostraron deficiencias en los aspectos gramaticales. A partir de esto, se sugiere que el dominio gramatical se evalúe con ítems acordes con el enfoque comunicativo, que muestren el uso de la lengua en contexto y no con ítems de selección múltiple (Rocha y Cuadros, 2008).

En la figura 1 se observa que la D2 (inferencia idea principal), la D3 (inferencia idea global), la D5 (género ensayo) y la D6 (género presentación oral) presentan resultados más bajos, es decir, son más difíciles para los estudiantes, dado que se trata de habilidades más complejas. Los resultados evidencian que el comportamiento de estas dimensiones es adecuado y esperable, puesto que las dimensiones inferenciales a nivel de la idea principal de párrafo, inferencia de la idea global, realizar una presentación oral y escribir un ensayo son más difíciles que la identificación de información literal (Martínez, 2011). En consecuencia, los resultados confirman las propuestas teóricas y de diseño que sustentan este prototipo de prueba en relación con los niveles de complejidad de las dimensiones comprensivas y productivas, por lo que se corrobora la validez de contenido del prototipo de prueba realizada por los expertos, dado que los ítems están midiendo las dimensiones que dicen medir (nivel B2 en EFA).

En efecto, se sugiere el uso de instrumentos que permitan recoger datos de manera ordenada y que faciliten el trabajo de quienes participen como jueces expertos. Como señalan Alderson *et al.* (1998), para evaluar la validez de contenido se debe elaborar documentos claros que permitan interpretar de manera adecuada las opiniones de los expertos en función de los objetivos de la investigación.

4.2 Validez de contenido de los aspectos tecnológicos de la prueba

En la tabla 4 se presentan los resultados de la otra validación de contenido de la prueba, referida a la evaluación del instrumento en su modalidad de aplicación en línea. Los ítems del cuestionario de escala Likert se elaboraron bajo las directrices metodológicas de CALT (Noijons, 1994; Chapelle y Douglas, 2006).

Dimensión	De acuerdo		Muy de acuerdo	
	n	%	n	%
Evaluativa	0	0,0	10	100,0
	1	10,0	9	90,0
	1	10,0	9	90,0
	2	20,0	8	80,0
	2	20,0	8	80,0
	2	20,0	8	80,0
Técnica	1	10,0	9	90,0
	0	0,0	10	100,0
	1	10,0	9	90,0
	0	0,0	10	100,0
	1	10,0	9	90,0
	0	0,0	10	100,0
	2	20,0	8	80,0
	1	10,0	9	90,0
	1	10,0	9	90,0
	1	10,0	9	90,0
Promedio		10,0		90,0

Ítems

1. La prueba cumple con el objetivo de evaluar el nivel de dominio B2 en EFA
 2. Los materiales (textos, audios, tareas) de la prueba evalúan el español en contextos auténticos, acorde con el enfoque comunicativo de enseñanza y evaluación de segundas lenguas
 3. Los diferentes tipos de ítems de la prueba son adecuados para evaluar las distintas habilidades lingüísticas acordes con el nivel B2 de EFA
 4. Considera adecuada la distribución de las secciones de la prueba: comprensión lectora, auditiva, competencia lingüística, expresión e interacción escrita y oral
 5. Considera adecuada la duración de la prueba
-
1. Es fácil la navegación en la página (botones visibles y simples)
 2. Es simple acceder y finalizar la sesión de la prueba
 3. La prueba tiene una conectividad segura (requiere de usuario y clave, de fecha y hora de aplicación)
 4. La prueba en línea evita elementos distractores (animaciones, figuras, exceso de botones, etc.)
 5. La prueba tiene una velocidad de carga adecuada
 6. Las instrucciones de la prueba se presentan con claridad
 7. El volumen de los audios es adecuado
 8. Las preguntas se presentan de manera ordenada y con un tamaño de letra adecuado
 9. Le parece interesante que se presenten todas las preguntas de una sección en una sola página
 10. En general, considera que esta prueba en línea está implementada en un formato simple

Tabla 4. Resultados validación tecnológica

Se observa en la tabla 4 que la primera dimensión del prototipo de prueba tuvo una evaluación positiva por parte de los expertos, puesto que sus respuestas se situaron en los niveles *de acuerdo* y *muy de acuerdo* (4 y 5 de la escala). De hecho, la evaluación mayoritaria está centrada en el valor máximo de la escala (*muy de acuerdo*). Por ejemplo, en el primer ítem se aprecia que el prototipo de prueba en línea está cumpliendo en un 100% con los propósitos de evaluar el español con fines específicos académicos en un nivel B2. El

segundo ítem muestra que un 90% de los jueces señala estar *muy de acuerdo* en que los materiales evalúan el español en un contexto comunicativo auténtico. Del mismo modo, un 90% marcó la opción *muy de acuerdo* cuando se les consultó si los diferentes tipos de ítems eran los adecuados para evaluar las diferentes habilidades lingüísticas; mientras que un 80% de los jueces está *muy de acuerdo* con la distribución de las secciones del prototipo de prueba y con su duración.

Los resultados de la segunda dimensión del prototipo de prueba indican que, efectivamente, se trata de un instrumento que cumple con lo exigido por la metodología de CALT. En el primer ítem, un 80% de los jueces está *muy de acuerdo* en que el ambiente donde se encuentra la prueba es de fácil navegación (ver ejemplo de la implementación en línea, en anexo 4), lo cual es esencial en una prueba de segundas lenguas, dado que el estudiante puede desplazarse sin problemas en la página y no requiere ayuda del profesor. En el segundo ítem, un 90% de los expertos señaló que el acceso y el cierre de la sesión en el prototipo de prueba es una tarea simple. En cuanto a la seguridad de la prueba, requisito fundamental en la evaluación de una segunda lengua, los 10 jueces indicaron estar *muy de acuerdo* en que el prototipo de prueba es seguro, puesto que se necesita que un profesor/administrador registre a los estudiantes, les genere un usuario, una clave y les asigne la fecha y la duración. Los resultados del ítem 4 indican que el 90% de los jueces está *muy de acuerdo* con que el prototipo de prueba evita elementos externos que distraigan a los estudiantes. En el ítem 5, el 100% de los jueces señaló que la prueba tiene una velocidad de carga adecuada. En el ítem 6, un 90% de los jueces está *muy de acuerdo* en que las instrucciones del prototipo de prueba son adecuadas, es decir, son comprensibles y suficientes para quienes la rinden. Los resultados del ítem 7 muestran que el volumen de los audios de la prueba es adecuado. Asimismo, en el ítem 8, un 80% está *muy de acuerdo* con que las preguntas de la prueba se presentan de manera ordenada y con un tamaño de letra adecuado. Para el 90% de los jueces es un aporte relevante el hecho de que las preguntas de cada sección de este prototipo de prueba estén en una sola página, lo cual beneficia

al estudiante porque disminuye la ansiedad y el tiempo de navegación. Asimismo, el 90% de los evaluadores consideró que el prototipo de prueba EFA en línea de nivel B2 está implementado en un formato de fácil navegación y uso, lo que se condice con lo señalado por Chappelle y Douglas (2006) en cuanto a los aportes de CALT a la evaluación en una segunda lengua.

Ítems

1. Me sentí cómodo (a) respondiendo la prueba en computador
2. Me parece interesante que la prueba esté disponible en computador
3. El ambiente de la prueba es fácil de usar
4. La prueba me pareció fácil de responder
5. Las instrucciones de la prueba son claras
6. Los textos y audios de la prueba son interesantes
7. Fue fácil comprender los textos de la prueba
8. Fue fácil comprender los audios de la prueba
9. Encontré fácil la sección de gramática
10. Me pareció fácil la tarea de escritura
11. Me pareció fácil la tarea oral
12. El tiempo de duración de la prueba es adecuado

4.3 Validez de apariencia

Los estudiantes extranjeros de intercambio universitario rindieron la prueba EFA, en línea, y luego respondieron un cuestionario de escala Likert de 5 puntos. De esta manera se pudo conocer la opinión de los alumnos frente al instrumento.

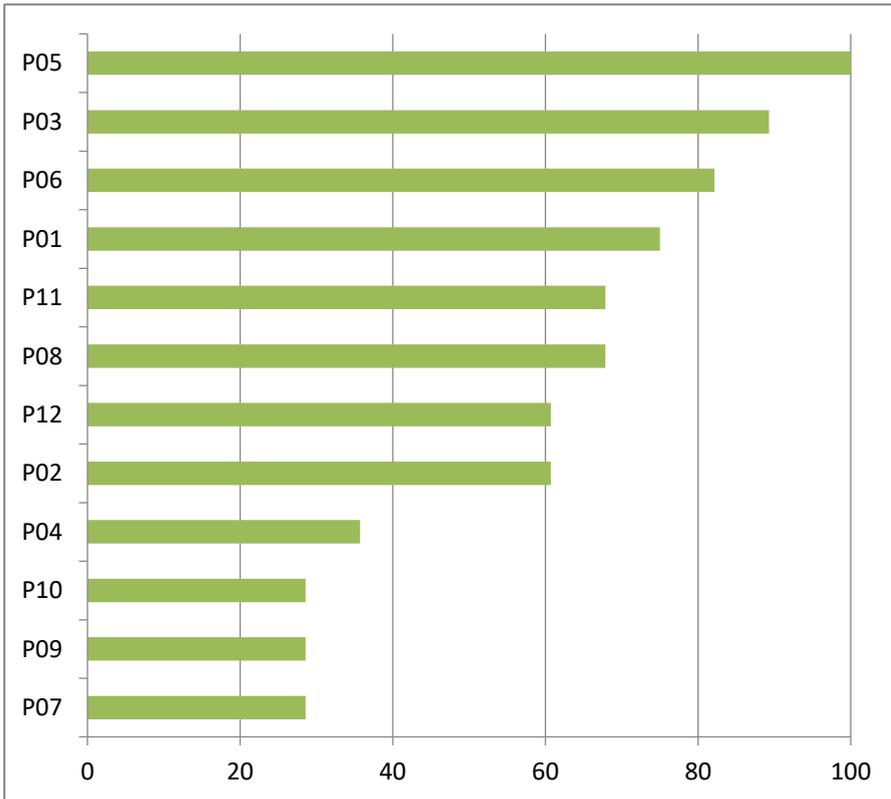


Figura 2. Resultados validez de apariencia

Se puede apreciar en la figura 2 que todos los estudiantes están muy de acuerdo en que las instrucciones de la prueba son claras, es decir, cada uno comprendió lo que debía hacer en cada sección (p5). Sin duda, este es un asunto muy relevante en todo instrumento evaluativo (Alderson *et al.* 1998; Martínez, 2011). Otros aspectos bien calificados por la

muestra se refieren a la facilidad de uso de la plataforma (p3) y a las temáticas que tratan los textos orales y escritos (p6), lo cual evidencia autenticidad en el uso de la lengua (Figueras y Puig, 2013).

Se observa un bajo nivel de acuerdo en las preguntas 4, 10, 9 y 7. Esto indica que para los estudiantes la prueba no les resultó fácil de responder, puesto que sus niveles de competencia no están acordes, en algunos casos, con el nivel B2, esperable en el contexto de un intercambio académico (Pastor, 2010).

En síntesis, los aspectos mejor valorados por los estudiantes extranjeros demuestran que se trata de un prototipo de prueba válido, dado que no presenta dificultades en la ejecución. Las mayores dificultades se deben a la falta de dominio lingüístico por parte de los alumnos de intercambio. Por lo mismo, se necesita contar con instrumentos validados que permitan diagnosticar con precisión los niveles de competencia en EFA y emprender acciones de apoyo que ayuden a los estudiantes en su inserción académica en universidades de habla hispana.

5. CONCLUSIONES

El aumento de los intercambios académicos en universidades de habla hispana ha evidenciado, por un lado, la necesidad de medir el nivel de español con el cual se integran los estudiantes extranjeros (cuya lengua materna no es el español) a las aulas y, por otro, la falta de instrumentos para evaluar el español como lengua extranjera en contextos académicos. Esto se debe, principalmente, a que las universidades no exigen algún tipo de certificado a los estudiantes o no han delimitado criterios claros respecto del nivel mínimo que se debería solicitar para garantizar un desempeño académico óptimo (Ainciburu y Rodríguez, 2015). En este sentido, apremia contar con instrumentos validados que definan con precisión los niveles de competencia en EFA.

Con los resultados obtenidos se ha dado cumplimiento a los dos objetivos planteados en este estudio, puesto que se

ha evidenciado la validez de contenido y de apariencia del prototipo de prueba EFA, en línea, de nivel B2.

Los datos de la validez de contenido evidencian un alto grado de coincidencia entre los jueces expertos, lo cual significa que los ítems evalúan las dimensiones definidas para el nivel de competencia B2 en EFA. Esto se corroboró con el desempeño de los estudiantes extranjeros de intercambio universitario en el prototipo de prueba. No obstante, se necesita aumentar el número de sujetos, así como también el número de ítems. Se debe tener en cuenta que la validación de una prueba no es algo que finaliza cuando esta se ha administrado y evaluado, sino que se trata de un proceso que se realiza continuamente (Bordón y Liskin-Gasparro, 2007), y es hacia allá donde se necesita avanzar en las pruebas de dominio en EFA.

La validación de contenido, en cuanto a los aspectos tecnológicos de la prueba, muestra que se trata de un instrumento adecuado, puesto que evalúa el español en un contexto académico comunicativo y en un entorno seguro y de fácil uso. En este sentido, la validación desde una perspectiva tecnológica proporciona información valiosa respecto de lo que significa evaluar de manera efectiva, rápida y con altos niveles de seguridad. Además, plantear una prueba de EFA, en línea, implica diseñar una prueba de lengua acorde con el perfil de los estudiantes universitarios.

En cuanto a la validez de apariencia, es preciso señalar que este tipo de validez aporta información cualitativa y cuantitativa relevante respecto del comportamiento del instrumento frente a una muestra real de participantes a quienes se orienta la prueba. Esta validez se suele obviar en la mayoría de los análisis de pruebas de segundas lenguas o lenguas extranjeras por considerarse poco objetiva (Alderson *et al.* 1998), sin embargo, en esta investigación contribuyó con información valiosa que permitió reformular, en primera instancia, algunos ítems y tareas de la prueba que presentaban un alto nivel de acierto o una redacción ambigua, por lo que se sugiere considerar la participación de hablantes nativos en pilotajes de futuras pruebas de EFA.

Una limitación de este estudio es el reducido número de estudiantes de español como lengua extranjera que

participaron en la aplicación del instrumento, teniendo en cuenta que para llevar a cabo procesos de validación de una prueba se requieren muestras más numerosas. No obstante, a partir de estos resultados se precisan nuevos estudios que indaguen sobre: el nivel de español que presentan los estudiantes extranjeros que realizan intercambios universitarios en universidades de habla hispana y las necesidades lingüísticas del alumnado para lograr una adecuada integración y un rendimiento académico óptimo.

En cuanto al proceso de validación de un examen se necesita evaluar el instrumento desde distintas perspectivas. En este sentido, en la evaluación del EFA se requiere de estudios que aporten sobre la validez de respuesta, validez externa y de constructo sobre las pruebas, de modo que existan medidas externas con las cuales se pueda comparar el nivel de español de los estudiantes extranjeros de intercambio universitario.

REFERENCIAS BIBLIOGRÁFICAS

Ainciburu, M., y Rodríguez, P. (2015). La elección de géneros textuales en los exámenes de certificación de lenguas académicas. Las necesidades del alumno universitario y la evaluación implícita de competencias no lingüísticas. En Morimoto, Y., Pavón, M., y Santamaría, R. (eds.): *La enseñanza de ELE centrada en el alumno*, XXV Congreso Internacional ASELE, Madrid: Universidad Carlos III, 87-96.

Alderson, Ch., Clapham, C., y Wall, D. (1998). *Exámenes de idiomas*, Madrid: Cambridge University Press.

Bachman, L. (1990). *Fundamental Considerations in Language Testing*, Oxford: Oxford University Press.

Bachman, L., y Palmer, A. (1996). *Language Testing and Practice*, Oxford: Oxford University Press.

- Bordón, T. (2015). La evaluación de segundas lenguas (L2). Balance y perspectivas. *Revista Internacional de Lenguas Extranjeras*, 4, 9-30.
- Bordón, T. (2006). *La evaluación de la lengua en el marco de E/L2: Bases y procedimientos*, Madrid: Arco Libros.
- Bordón, T., y Liskin-Gasparro, J. (2007). "Evaluación". En: Lacorte, M. (ed.): *Lingüística aplicada del español*, Madrid: Arco/Libros, 211-252.
- Buck, G. (2009). *Assessing Listening*, Cambridge: Cambridge University Press.
- Cabero, J., y Llorente, M. (2013). La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información y comunicación (TIC), *Eduweb*, 7, (2), 11-22.
- Consejo de Europa. (2002) *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*, Madrid: Anaya.
- Cronbach, L., y Meehl, P. (1955). Construct Validity in Psychological Tests, *Psychological Bulletin*, 52 (4), 281-302.
- Chapelle, C. (2017). Conceptions of validity. En Fulcher, G y Davidson, F (eds.): *The Routledge Handbook of Language Testing*: Routledge Taylor & Francis Group, 21-33.
- Chapelle, C., y Douglas, D. (2006). *Assessing language through computer technology*, United Kingdom: Cambridge University Press.

- Chapelle, C., y Voss, E. (2016), 20 years of technology and language assessment in Language Learning & Technology, *Language Learning and Technology*, 20 (2), 116-128.
- Escobar, J., y Cuervo, Á. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización, *Avances en medición*, 6 (1), 27-36.
- Ferreira, A., Vine, A., y Elejalde, J. (2015). Diseño y evaluación de una aplicación tecnológica para la enseñanza del español como lengua extranjera, *Onomázein*, 31(1), 145-166.
- Figueras, N., y Puig, F. (2013). *Pautas para la evaluación del español como lengua extranjera*, Madrid: Edinumen.
- Fulcher, G. (2003). *Testing second language speaking*, New York: Pearson Education.
- García, B., Martín, S., y Masid, O. (2016). Evaluación de la comprensión oral en ELE en el contexto académico, ponencia presentada en *34 Congreso Internacional de AESLA*, Alicante.
- García, J., y Navarro, C. (2008). Automatizar los exámenes del Diploma de Español como Lengua Extranjera (DELE): ¿Utopía o realidad?", *Revista de Lingüística Teórica y Aplicada*, 46 (2), 81-93.
- Jouini, K. (2005). Estrategias inferenciales en la comprensión lectora, *Glosas didácticas*, 13, 95-114.
- Juan, O. (2015). Efectividad de la mensajería externa en la adquisición de la competencia comunicativa en un modelo de aprendizaje autónomo y a distancia en español como lengua extranjera, *Revista de Lingüística Teórica y Aplicada*, 53 (2), 121-143.

- Juan, O. (2009). Web 2.0, comunicación y Material Didáctico Digital para el aprendizaje del español: El Aula Virtual de Español el Instituto Cervantes y su actualización, *Revista de Lingüística Teórica y Aplicada*, 47 (2), 13-34.
- Kane, M. (2006). Validation. En Brennan, R. (ed.): *Educational Measurement*, Westport, American Council on Education: Praeger, 17-64.
- Kane, M. (1992). An Argument-Based Approach to Validation, *Psychological Bulletin*, 112 (3), 527-535.
- Llorián, S. (2012). Los retos de la certificación del español con fines específicos, *Revista de Nebrija de Lingüística Aplicada*, 12, 1-34. <http://www.nebrija.com/revista-linguistica/los-retos-de-la-certificacion-del-espa%C3%B1ol-con-fines-especificos>
- Martín, S. (2010). La elaboración de textos y preguntas para evaluar la comprensión oral. En *Actas del XXI Congreso Internacional ASELE*, 557-569.
- Martínez, A., Biedma, A., Sánchez, A., Chamorro, M., y Sánchez, S. (2015). Descripción y análisis del examen online de acreditación de dominio de español de los niveles B1 y B2: el eLADE, *Certiuni Journal*, 1, 52-66.
- Martínez, A. (2011). *La evaluación de las lenguas: garantías y limitaciones*, Granada: Ediciones Mágina.
- Mendoza, A. (2015). La validez en los exámenes de alto impacto: Un enfoque desde la lógica argumentativa, *Perfiles educativos*, 37(149), 169-186.
- Ministerio de Educación. (2016). *Situación de la matrícula de extranjeros en educación superior en Chile: Datos 2015*, Santiago: Servicio de Información de Educación Superior.

- Pastor, S., y Pandor, J. (2017). El choque cultural académico del alumnado estadounidense en la universidad española, *Revista de Lingüística Teórica y Aplicada*, 55 (2), 13-38.
- Pastor, S. (2010). Enseñanza de español con fines profesionales y académicos y aprendizaje por contenidos en contexto universitario, *Testi e Linguaggi*, 4, 71-88.
- Noijons, J. (1994). Testing computer assisted language testing: Towards a checklist for CALT, *CALICO Journal*, 12(1), 37-58.
- Quevedo, Á. (2017). El español con fines académicos: progreso de un estudio sobre necesidades lingüísticas de estudiantes en programas de movilidad, *Revista Nebrija de Lingüística Aplicada*, 22, 1-7.
<http://www.nebrija.com/revista-linguistica/el-espa%C3%B1ol-con-fines-academicos-progreso-de-un-estudio-sobre-necesidades-linguisticas-de-estudiantes-en-programas-de-movilidad>
- Ramak, M., y Rahmattalab, S. (2017). Computer Assisted Language Testing Benefits, Constraints and Future Directions, *Research Journal of English Language and Literature*, 5 (2), 336-342.
- Regueiro, M. y Sáez, D. (2013) *El español académico*, Madrid: Arco Libros.
- Rocha, C., y Cuadros, R. (2008). ¿Es posible evaluar objetivamente contenidos gramaticales en producciones escritas? En: Pastor, S., y Roca, S. (eds.): *La evaluación en el aprendizaje y la enseñanza del español como lengua extranjera/segunda lengua*, Alicante: Universidad de Alicante, 179-184.

- Sireci, S. (1998). The construct of content validity, *Social Indicators Research*, 45, 83-117.
- Vine, A., y Ferreira, A. (2016). Propuesta de un modelo para una prueba con fines específicos académicos en ELE, *Literatura y Lingüística*, 33, 369-390.
- Vine, A. (2015). *Criterios evaluativos para el diseño de un prototipo de una prueba de Español como Lengua Extranjera con fines específicos Académicos (EFA)*. Tesis de Doctorado, Universidad de Concepción.
- Winke, P., e Isbell, D. (2017). Computer-Assisted Language Assessment, *Language, Education and Technology*, 2, 1-13.

ANEXO 1: CUESTIONARIO ABREVIADO SOBRE LOS ÍTEMS DEL PROTOTIPO PRUEBA EFA B2

Dimensión I: Identifica información literal en textos especializados

Sección comprensión de lectura	EVALÚA DIMENSIÓN	INDICADORES				
		1	2	3	4	5
Preguntas	SI/NO (observación)					
<i>Texto I</i>						
1. De acuerdo con el texto, la ingeniería realiza aportes relevantes al desarrollo de un país.						
2. Según el texto, la formación que reciben los ingenieros les impide encontrar trabajo.						

Dimensión II: Infiere las ideas principales de los párrafos

Sección comprensión de lectura	EVALÚA DIMENSIÓN	INDICADORES				
		1	2	3	4	5
Preguntas	SI/NO (observación)					
<i>Texto III</i>						
3. La situación actual de la biodiversidad indica que						

4. En el futuro se espera que haya...						
---------------------------------------	--	--	--	--	--	--

Dimensión III: Infiere la idea global del texto

Sección comprensión de lectura	EVALÚA DIMENSIÓN	INDICADORES				
Preguntas	SI/NO (observación)	1	2	3	4	5
<i>Texto I</i>						
5. A partir de la lectura del texto se puede inferir que la sociedad requiere de egresados vinculados al ámbito laboral.						

Dimensión IV: Competencia gramatical B2

	EVALÚA DIMENSIÓN	INDICADORES				
Sección competencia lingüística	SI/NO (observación)	1	2	3	4	5
Texto para completar con las formas gramaticales adecuadas.						

Dimensión V: Género académico ensayo

	EVALÚA DIMENSIÓN	INDICADORES				
Tareas	SI/NO (observación)	1	2	3	4	5
<i>Tarea de expresión e interacción escrita:</i> Ensayo						

Dimensión VI: Género académico exposición oral

	EVALÚA DIMENSIÓN	INDICADORES				
Tareas	SI/NO (observación)	1	2	3	4	5
<i>Tarea de expresión e interacción oral:</i> Exposición oral						

ANEXO 2: CUESTIONARIO MODALIDAD APLICACIÓN EN LÍNEA

D	Ítems	1	2	3	4	5
Dimensión Evaluativa	1.La prueba cumple con el objetivo de evaluar el nivel de dominio B2 en EFA					
	2. Los materiales (textos, audios, tareas) de la prueba evalúan el español en contextos auténticos, acorde con el enfoque comunicativo de enseñanza y evaluación de segundas lenguas					
	3. Los diferentes tipos de ítems de la prueba son adecuados para evaluar las distintas habilidades lingüísticas acordes con el nivel B2 de EFA					
	4. Considera adecuada la distribución de las secciones de la prueba: comprensión lectora, auditiva, competencia lingüística, expresión e interacción escrita y oral					
	5. Considera adecuada la duración de la prueba					
Dimensión técnica	1.Es fácil la navegación en la página (botones visibles y simples)					
	2. Es simple acceder y finalizar la sesión de la prueba					
	3.La prueba tiene una conectividad segura (requiere de usuario y clave, de fecha y hora de aplicación)					
	4. La prueba en línea evita elementos distractores (animaciones, figuras, exceso de botones, etc.)					

5. La prueba tiene una velocidad de carga adecuada					
6. Las instrucciones de la prueba se presentan con claridad					
7. El volumen de los audios es adecuado					
8. Las preguntas se presentan de manera ordenada y con un tamaño de letra adecuado					
9. Le parece interesante que se presenten todas las preguntas de una sección en una sola página					
10. En general, considera que esta prueba en línea está implementada en un formato simple					

ANEXO 3: CUESTIONARIO VALIDEZ DE APARIENCIA APLICADO A ALUMNOS DE ESPAÑOL COMO LENGUA EXTRANJERA

I. Marca el número que mejor represente tu opinión en relación con la pregunta

1: Muy en desacuerdo 2: En desacuerdo 3: Indiferente 4: De acuerdo 5: Muy de acuerdo

Preguntas	1	2	3	4	5
1. Me sentí cómodo (a) respondiendo la prueba en computador					
2. Me parece interesante que la prueba esté disponible en computador					
3. El ambiente de la prueba es fácil de usar					

4. La prueba me pareció fácil de responder					
5. Las instrucciones de la prueba son claras					
6. Los textos y audios de la prueba son interesantes					
7. Fue fácil comprender los textos de la prueba					
8. Fue fácil comprender los audios de la prueba					
9. Encontré fácil la sección de gramática					
10. Me pareció fácil la tarea de escritura					
11. Me pareció fácil la tarea oral					
12. El tiempo de duración de la prueba es adecuado					

ANEXO 4: EJEMPLO DE LA IMPLEMENTACIÓN EN LÍNEA DEL PROTOTIPO DE PRUEBA EFA B2



Audio 3: Escucha el audio y selecciona la opción correcta (a, b, c, d)

Según el audio, el petróleo es un combustible...

El Centro de Biotecnología de la Universidad de Concepción está interesado en...

El Centro de Biotecnología puede llevar a cabo su investigación porque la región donde se encuentra cuenta con...

De acuerdo con el audio, las microalgas...

¿Cuál es el tema principal del audio?

Sección IV: Expresión e interacción escrita

La sección de expresión e interacción escrita consta de dos opciones. Solo tienes que elegir UNA opción y escribir un ensayo de 250 palabras (20-25 líneas)

Opción 1

Una alternativa a la formación académica tradicional es tomar cursos a distancia (por Internet), modalidad que permite un acceso real al conocimiento de personas que antes no podían hacerlo, ya sea por razones económicas o de tiempo. La modalidad de aprendizaje a distancia, cada vez va tomando más fuerza, pero aún queda el escepticismo de algunos que dudan si es efectiva o no.

En relación con este tema, escribe un ensayo sobre la formación académica a través de cursos a distancia por Internet.

En el ensayo tienes que incluir:

- Escribir un título
- Introducir el tema
- Establecer una postura respecto del tema
- Mencionar las ventajas y desventajas de la formación a través de internet
- Concluir