



VALIDEZ EN LAS EVALUACIONES DE INGLÉS PARA LAS CIENCIAS ECONÓMICAS: AMENAZAS Y CÓMO CONTROLARLAS

ASSESSMENT VALIDITY IN ENGLISH FOR THE ECONOMIC SCIENCES: THREATS AND HOW TO CONTROL THEM

Autores

Luciani, María de las Mercedes- Chavez, Magdalena- Di Giuseppe, Maria Salome- Lauría, Carla Mariana

Email

mluciani@fce.unl.edu.ar

Eje Temático

Educación en las Ciencias Económicas

Modalidad

Ensayos, producciones y/o comunicaciones

Palabras claves: validez, evaluación, comprensión oral y escrita

Resumen

Este trabajo se centra en el concepto de validez en la evaluación, analizado particularmente con respecto a la asignatura *Inglés Técnico* de las carreras de Contador Público Nacional, Licenciatura en Economía y Licenciatura en Administración de la Facultad de Ciencias Económicas de la UNL. En primer lugar, se trata el significado de validez y su impacto en el diseño e implementación de exámenes para acreditar esta asignatura. Se reflexiona sobre los diferentes factores que atentan contra la validez en las evaluaciones y las acciones y recursos que es posible implementar para controlar los elementos disruptivos en el proceso de diseño de los exámenes, en este caso en particular, la evaluación de dos macro-destrezas, la comprensión oral y escrita en la



asignatura *Inglés Técnico*. Para concluir, se mencionan las acciones que se llevan a cabo a fin de evaluar las evaluaciones para introducir mejoras y/o que fundamenten la continuidad de las fortalezas existentes en el proceso de evaluación, a fin de que este contribuya positivamente al aprendizaje y la enseñanza del inglés para las ciencias económicas en nuestra Facultad.

Introducción

El diseño de exámenes que gocen de validez para la asignatura *Inglés Técnico*, que es parte del Plan de Estudio de las tres carreras, Contador Público Nacional, Licenciatura en Administración y Licenciatura en Economía de la Facultad de Ciencias Económicas, UNL, puede parecer una tarea simple. Sin embargo, es un verdadero desafío minimizar las posibles amenazas a la validez de los mismos. Se han determinado los niveles de proficiencia en la lengua esperados para esta asignatura de acuerdo a un conjunto de niveles de logro propuestos por el Marco Común Europeo de Referencia para las Lenguas (MCER), el cual en este contexto se utiliza como estándar de logro y guía para el diseño e interpretación de los resultados de los correspondientes exámenes. El propósito de este trabajo es reflexionar sobre dos constructos que forman parte del concepto de validez en la evaluación: validez de contenido y dificultad de las tareas en los instrumentos utilizados en la evaluación a fin de identificar potenciales amenazas a la validez de constructo relacionada con la evaluación de macro-habilidades específicas, comprensión oral y escrita, en un contexto de aplicación específico, la asignatura *Inglés Técnico* en la Facultad de Ciencias Económicas de la UNL.

Contexto de aplicación de los exámenes

La asignatura *Inglés Técnico* es una asignatura obligatoria que forma parte del Plan de Estudio de las carreras de Contador Público Nacional, Licenciatura en Economía y Licenciatura en Administración de la Facultad de Ciencias Económicas de la UNL. Los estudiantes tienen que aprobar las dos asignaturas del Ciclo Inicial de Inglés, Idioma Extranjero-inglés I y II, para poder cursar *Inglés Técnico*, según el sistema de correlatividades. El Programa de esta asignatura se basa en el *Marco Común Europeo de Referencia para las Lenguas* para la determinación de sus niveles de logro, que difieren según sea la macro-habilidad en cuestión (Escucha B1, Lectura B2-, Escritura B1 y Oralidad B1).



¿Qué significan estos niveles de logro? El Marco Común Europeo de Referencia para las lenguas es un documento elaborado por el Consejo Europeo y constituye una base para la elaboración de guías curriculares, exámenes, libros de texto y también sirve para medir el nivel de comprensión y expresión oral y escrita en una lengua determinada. Adopta un enfoque orientado a la acción describe de un modo comprensivo lo que un aprendiente de lengua necesita saber a fin de usar una lengua para la comunicación y qué conocimiento y habilidades deben desarrollar para actuar efectivamente. utiliza conceptos clave como *conocimiento procedimental*, *conocimiento declarativo* y *competencia*. Esta última se entiende como el conocimiento mínimo y necesario que se requiere para llevar a cabo tareas específicas. El MCER propone una escala de seis niveles de logro que sirven de referencia para el aprendizaje y permiten la medición del progreso de los aprendientes de una lengua en distintas etapas de su aprendizaje.

Usuario competente	C2	Es capaz de comprender con facilidad prácticamente todo lo que oye o lee. Sabe reconstruir la información y los argumentos procedentes de diversas fuentes, ya sean en lengua hablada o escrita, y presentarlos de manera coherente y resumida. Puede expresarse espontáneamente, con gran fluidez y con un grado de precisión que le permite diferenciar pequeños matices de significado incluso en situaciones de mayor complejidad.
	C1	Es capaz de comprender una amplia variedad de textos extensos y con cierto nivel de exigencia, así como reconocer en ellos sentidos implícitos. Sabe expresarse de forma fluida y espontánea sin muestras muy evidentes de esfuerzo para encontrar la expresión adecuada. Puede hacer un uso flexible y efectivo del idioma para fines sociales, académicos y profesionales. Puede producir textos claros, bien estructurados y detallados sobre temas de cierta complejidad, mostrando un uso correcto de los mecanismos de organización, articulación y cohesión del texto.
Usuario independiente	B2	Es capaz de entender las ideas principales de textos complejos que traten de temas tanto concretos como abstractos, incluso si son de carácter técnico, siempre que estén dentro de su campo de especialización.



		<p>Puede relacionarse con hablantes nativos con un grado suficiente de fluidez y naturalidad, de modo que la comunicación se realice sin esfuerzo por parte de los interlocutores.</p> <p>Puede producir textos claros y detallados sobre temas diversos, así como defender un punto de vista sobre temas generales, indicando los pros y los contras de las distintas opciones.</p>
	B1	<p>Es capaz de comprender los puntos principales de textos claros y en lengua estándar si tratan sobre cuestiones que le son conocidas, ya sea en situaciones de trabajo, de estudio o de ocio.</p> <p>Sabe desenvolverse en la mayor parte de las situaciones que pueden surgir durante un viaje por zonas donde se utiliza la lengua.</p> <p>Es capaz de producir textos sencillos y coherentes sobre temas que le son familiares o en los que tiene un interés personal.</p> <p>Puede describir experiencias, acontecimientos, deseos y aspiraciones, así como justificar brevemente sus opiniones o explicar sus planes.</p>
Usuario básico	A2	<p>Es capaz de comprender frases y expresiones de uso frecuente relacionadas con áreas de experiencia que le son especialmente relevantes (información básica sobre sí mismo y su familia, compras, lugares de interés, ocupaciones, etc.).</p> <p>Sabe comunicarse a la hora de llevar a cabo tareas simples y cotidianas que no requieran más que intercambios sencillos y directos de información sobre cuestiones que le son conocidas o habituales.</p> <p>Sabe describir en términos sencillos aspectos de su pasado y su entorno, así como cuestiones relacionadas con sus necesidades inmediatas.</p>
	A1	<p>Es capaz de comprender y utilizar expresiones cotidianas de uso muy frecuente, así como, frases sencillas destinadas a satisfacer necesidades de tipo inmediato.</p> <p>Puede presentarse a sí mismo y a otros, pedir y dar información personal básica sobre su domicilio, sus pertenencias y las personas que conoce.</p> <p>Puede relacionarse de forma elemental siempre que su interlocutor hable despacio y con claridad y esté dispuesto a cooperar</p>

Tabla 1. Niveles comunes de referencia: escala global. Marco Común Europeo de Referencia para las lenguas. (p.26)



La decisión de establecer un punto de corte entre B1 y B2 para la macro-habilidad de la comprensión lectora y B1 para las otras habilidades surge de la consideración de los niveles de logro de las asignaturas previas y de las características y necesidades de los estudiantes. Esto permite manejar un grado de sutileza de las características de los niveles de logro que ayuda a establecer objetivos de aprendizaje alcanzables para asegurar la necesaria legitimidad en la evaluación. Para acreditar esta asignatura, los estudiantes deben aprobar dos parciales que cubren las cuatro macro-habilidades a evaluar y que se mencionaron previamente.

¿Qué se entiende por validez?

La validez es un tema central en toda evaluación para los encargados del diseño de exámenes ya que su ausencia tiene un serio impacto en todos aquellos involucrados en una evaluación. Tradicionalmente, se definía la validez en relación al examen como en “[v]alidity is about the test: A test is valid if it measures what it is supposed to be measuring (Kelley, T.L., 1927)¹, pero más reciente mente Messick (1980) afirmó:

"Validity is about inferences: A test is valid if the results support the decisions we want to make – i.e. the inferences we make from the test scores must be valid." (S. Messick, 1980)²

Posteriormente, S. Messick (1982, p.13) propone otra definición de validez:

"an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and

¹ “la validez se refiere al examen: un examen es válido si mide lo que se supone que debe medir” (Kelley, T.L., 1927) (Traducción propia).

² “La validez se refiere a las inferencias: Un examen es válido si sus resultados respaldan las decisiones que deseamos tomar – es decir, las inferencias que realizamos a partir de las calificaciones de los exámenes deben ser válidas” (S. Messick, 1980). (traducción propia)



appropriateness of inferences and actions based on test scores or other modes of assessment.”³

Tanto la medición válida como las inferencias precisas son el resultado esperado de cualquier evaluación. Teniendo en cuenta que la evaluación es una instancia situada, su validez está inextricablemente relacionada con el uso específico de la evaluación en un contexto particular, administrado a un grupo de estudiantes específico que esperan que sus conocimientos y competencias sean acreditados.

La evaluación debería constituir una base confiable para obtener inferencias válidas sobre el desarrollo del conocimiento y las competencias de los estudiantes evaluados. Por lo tanto, deberíamos tener la certeza de que el constructo que es el centro de la evaluación está claramente definido y que se corresponde con los objetivos del aprendizaje explicitados en el Programa de una asignatura dada. Otro aspecto de la validez, la de contenido, también demanda una cuidadosa atención. La validez de contenido responde a la pregunta: ¿“hasta qué punto los elementos del conocimiento, comprensión y/o habilidades” (Cambridge Approach to Assessment, p. 13) de un examen cubren los contenidos que se esperan desarrollar de acuerdo al programa de la asignatura? En el caso particular de *Inglés Técnico*, los contenidos que se deben cubrir se encuentran delineados en el Programa de la asignatura y en el material didáctico de la materia. Por lo tanto, se espera que toda evaluación de esta asignatura esté alineada con el contenido incluido y con las competencias delineadas en el programa y, de gran importancia también, con el modo de enseñanza que atraviesa esos contenidos y competencias.

Contenido del examen y dificultad de los ítems en relación con la validez

Uno podría pensar que la utilización del Común Marco Europeo de Referencia (MCER) como un punto de referencia para los niveles de logro podría resolver cualquier dificultad relacionada con el diseño de la evaluación. Sin embargo, el

³ “un juicio evaluativo integrador del grado en el cual la evidencia empírica y las razones teóricas respaldan la adecuación de las inferencias basadas en las calificaciones y otros modos de evaluación.”



Manual del Consejo Europeo que proporciona las directrices para vincular las evaluaciones de lengua al MCER establece:

Examinations vary in content and style, according to the needs of their context and the traditions of the pedagogic culture in which they are developed, so two examinations may both be “at Level B2” and yet differ considerably. Learners in two different contexts might gain very different scores on (a) an examination whose style and content they are familiar with and (b) an examination at the same level developed for a different context. [...], the fact that several examinations may be claimed to be “at Level B2” as a result of following procedures to link them to the CEFR, such as those suggested in this Manual, does not mean that those examinations are claimed to be exactly the same level.⁴

En el proceso de diseño de las evaluaciones para *Inglés Técnico* en nuestra facultad, se deben tener en cuenta ciertas consideraciones con respecto al contenido del examen y al diseño del instrumento.

1. Contenido del examen

Para que el contenido de un examen sea válido debería ser una muestra adecuada de los elementos más significativos del área, es decir, aprendizaje de una lengua extranjera, específicamente inglés para las ciencias económicas. Si este no fuera el caso, las inferencias sobre el rendimiento de los alumnos en esa área en particular carecerían de validez. Por esta razón, los exámenes deberían cubrir los temas incluidos en el programa de estudio para la asignatura y deberían enfocarse en los elementos centrales, de modo que representen el

⁴ Los exámenes varían en contenido y estilo, de acuerdo a las necesidades de su contexto y a las tradiciones de la cultura pedagógica en la cual se desarrollan; por lo tanto, dos exámenes pueden ambos estar “a un nivel B2” y sin embargo difieren considerablemente. Los alumnos en dos contextos diferentes podrían obtener diferentes calificaciones en a) un examen cuyo estilo y contenido resulta familiar y (b) un examen al mismo nivel diseñado para un contexto diferente. [...] el hecho de que varios exámenes puedan considerarse como del mismo nivel, B2 por ejemplo, como resultado de cumplir con los procedimientos que los relacionan con el MCER, tales como los que se sugieren en este Manual, no significa que estén exactamente al mismo nivel. (traducción propia)



rango de contenidos relacionados tanto con el constructo a evaluar como con el nivel de logro correspondiente. De ese modo, se debe evitar caer en lo que se denomina *sub-representación del constructo*, que se entiende como la falta de inclusión en el examen de importantes dimensiones del constructo a evaluar. También se debe eludir lo que se conoce como *variables irrelevantes al constructo*, que es el resultado de la inclusión de elementos extraños al constructo a evaluar y que afectan los resultados del examen (The Cambridge Approach to Assessment, 2009, p.13). En consecuencia, se deben tomar todos los recaudos necesarios para diseñar ítems cuya estructura esté en línea con el dominio de contenido a evaluar.

Teniendo en cuenta lo expresado hasta aquí, podríamos preguntarnos qué área en particular del proceso de diseño de un examen es pasible de mostrar errores que afecten la validez del mismo. Sería conveniente, entonces, reflexionar acerca de las tareas que constituyen los ítems que conforman los instrumentos para la evaluación de la escucha y la comprensión lectora.

2. *Diseño de las tareas*

En nuestra vida cotidiana realizamos numerosas tareas y, en su mayoría, estas involucran la comunicación. Esa es la razón por la que forman parte de los programas de estudio, de las experiencias áulicas y, por supuesto, de los exámenes. Entonces, podemos hablar de tareas pedagógicas y tareas evaluativas. Las tareas comunicativas en una situación de aprendizaje o de examen requieren la activación de las competencias generales apropiadas y los aprendientes recurren a las competencias comunicativas (habilidades y conocimiento lingüísticos, sociolingüísticos y pragmáticos) para realizarlas con éxito (MCER, pp. 157-158). A pesar de que en nuestra vida cotidiana las tareas son omnipresentes, su concreción puede verse afectada por ciertas características y restricciones que dan como resultado diversos grados de dificultad de las mismas.

En las tareas de comprensión, tanto si se refieren a la escucha o la lectura, podemos mencionar un número de consideraciones que se relacionan directamente con la dificultad que impone una tarea dada, como lo explica el MCER:



2.1 Cantidad y calidad del input disponible para los examinados:

La primera etapa en el diseño de la tarea es la selección del texto (oral o escrito) de acuerdo con los niveles de logro esperados. Partiendo de una definición de texto que incluye la interpretación que realiza tanto el interlocutor como el lector, cualquier desequilibrio en la longitud y/o complejidad de la información afecta seriamente los resultados del examen.

2.2 Límite de tiempo de resolución:

El número de veces que un audio para una tarea de escucha es accesible para los alumnos y el límite de tiempo que se establece para llevar a cabo la lectura de un texto escrito influyen en el nivel de dificultad de la tarea que los examinados deben completar.

2.3 Factores lingüísticos:

El nivel de los recursos lingüísticos que poseen los examinados constituye un factor central cuando se debe determinar si una tarea es apropiada para ese contexto de evaluación. El dominio que los examinados posean en cuanto a sus recursos de lengua (complejidad y precisión de los recursos gramaticales y léxicos) así como la etapa de desarrollo de diferentes aspectos del lenguaje en uso (fluidez, coherencia, adecuación, etc.) facilitarán u obstaculizarán la concreción de una tarea dada.

2.4 Factores cognitivos:

La carga cognitiva que se impone a los examinados puede reducirse o incrementarse de acuerdo al grado de familiaridad con el tipo de tarea, el tema, el tipo de texto (ej. género discursivo), conocimiento del contexto y conocimiento sociocultural. También se deben tener en cuenta las demandas procesuales requeridas por parte de los examinados para realizar las operaciones cognitivas necesarias para la resolución de la tarea.

2.5 Calidad del rendimiento esperado:

La calidad de las respuestas o resolución de la tarea debería estar en un todo de acuerdo con el nivel de proficiencia de los examinados.



Recomendaciones para controlar amenazas a la validez

Tener en cuenta los factores y principios mencionados hasta aquí no garantiza que un examen goce de total validez. ¿Qué se puede hacer para minimizar las posibles amenazas a la validez? Primero, en el proceso de selección y adaptación de un texto, se debe realizar un detallado análisis de los ítems léxicos, construcciones gramaticales, características discursivas y tema del texto a utilizar para la evaluación de manera que sea posible acomodar su nivel de dificultad al nivel de logro esperado de los examinados y, así, minimizar la parcialidad o arbitrariedad debida a un sesgo del examen. Para lograr esto es posible recurrir a algunos recursos que pueden resultar de ayuda: los descriptores de cada nivel del MCER, el proyecto English Profile⁵ que ofrece una clasificación de los ítems léxicos y gramaticales de acuerdo a los niveles de logro del MCER, y el diccionario Cambridge Advanced Learner's Dictionary que también ofrece un significativo número de palabras clasificadas de acuerdo a los niveles del MCER. Segundo, el criterio más importante a satisfacer es la pertinencia de un ítem de evaluación para una situación de uso en particular. También son esenciales la familiaridad con el tipo de ítem y la apropiada demanda cognitiva. Se debe evitar una *inapropiada dificultad del constructo* en los ítems basados en un texto, ya que un mejor rendimiento debe estar relacionado con los elementos del constructo y no con una indebida dificultad en el diseño del ítem. Pero también debe evitarse la *irrelevante facilidad del constructo*, que surge de claves o pistas en la tarea que permiten a los examinados responder correctamente sin relación con el constructo evaluado.

Sin duda el contenido y la dificultad de la tarea no son los únicos factores que atentan contra la validez de un examen. También puede suceder que el contexto

⁵ El proyecto *English Profile* describe los aspectos del inglés se espera que típicamente se aprendan en cada nivel del MCER. Es útil para los docentes, autores de libros de texto y para los que diseñan exámenes, pues ofrece sugerencias acerca de los que es apropiado para cada nivel. Este es un proyecto colaborativo apoyado por el Consejo de Europa y tiene como objetivo recolectar datos de los estudiantes de inglés en todo el mundo a fin de detectar los usos gramaticales y léxicos para cada uno de los 6 niveles del MCER. Los resultados de esta investigación se conocen como *Reference Level Descriptors (RLDs)* y están accesibles en el sitio del Proyecto. <http://www.englishprofile.org/>



de administración del examen no ofrezca las condiciones edilicias y tecnológicas adecuadas para llevar a cabo un examen. Por ejemplo, las cualidades de los equipos de audio, un ambiente ruidoso, insuficiente espacio para que los examinados estén cómodamente sentados, etc. pueden contribuir al malestar o incomodidad de los participantes, con la consiguiente disminución de su rendimiento. Sin embargo, esto no sucede en nuestra Facultad, la cual ofrece condiciones apropiadas para la administración de exámenes de esta naturaleza. Además, las clases y el material del entorno virtual ofrecen suficientes instancias de práctica y las muestras de exámenes contribuyen a familiarizar a los estudiantes con el tipo de instrumentos que aparecerán en las evaluaciones.

A modo de reflexión final

En este trabajo se trató el concepto de validez y su importancia para el sistema de evaluación, especialmente aplicado a la asignatura *Inglés Técnico* para evaluar la comprensión oral y escrita. Se plantearon, además, los factores que pueden jugar en detrimento de la validez de la evaluación para esta asignatura, con el consiguiente efecto negativo en los resultados que obtienen los estudiantes. También se sugirieron recursos y procedimientos que pueden contribuir a la validez en el diseño de las evaluaciones.

Ahora bien, ¿qué acciones concretas se están llevando a cabo para asegurar el proceso de validación de la evaluación en esta asignatura? La asignatura *Inglés Técnico* forma parte del proyecto de investigación "Meta-evaluación de los aprendizajes universitarios como insumo para mejorar la calidad educativa en inglés y Matemática" (CAI+D 2016), el cual tiene como objetivo evaluar las evaluaciones para determinar las condiciones epistemológicas, didácticas, culturales, humanas, tecnológicas y logísticas necesarias para determinar qué modificaciones son necesarias en el propio proceso de evaluación y para asegurar una retroalimentación de la evaluación sobre los procesos de enseñanza y aprendizaje. En este contexto, una de las acciones que se encuentran en curso en este momento es el análisis de los ítems utilizados en las diferentes evaluaciones en esta asignatura para determinar su validez con respecto a los objetivos y los niveles de logro especificados en el Programa de la asignatura. Esto se lleva a cabo mediante grillas específicamente diseñadas para evaluar todos los criterios que se han desarrollado en este trabajo, comparar esta evaluación con los resultados de los exámenes para cada ítem y determinar fortalezas y debilidades de cada ítem. Se creará de este modo un



sistema de meta-evaluación continua que contribuya a la mejora de la evaluación y a los procesos de enseñanza y de aprendizaje.

Bibliografía

- ALTE. *Materials for the Guidance of Test Item Writers*. (1995, updated July 2005)
- Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Strasbourg: Language Policy Division. Council of Europe.
- Marco Común Europeo de Referencia para las Lenguas*. 2002 Instituto Cervantes para la traducción en español. <http://cvc.cervantes.es/obref/marco>
- Issacs, T., Zara, C., and Herbert, G. with Coombs, S.J. and Smith, C. (2013) *Key Concepts in Educational Assessment*. Los Angeles, London, New Delhi, Singapore and Washington: Sage.
- Kelley, L. (1927) *Interpretation of educational measurements*. New York: Macmillan, 1927.
- Messick, S. (1989) "Validity". In Linn, R.L. (Ed.) *Educational Measurement* (3rd ed. p. 13). Old Tappan, NJ: Macmillan.
- Messick, S. *Test validity and the Ethics of Assessment*. *American Psychologist*. Vol 35(11), Nov. 1980
- The Cambridge Approach to Assessment*. Principles for designing, administering and evaluating assessment. January 2009-revised april 2017.
- Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (CEFR). (2009) Strasbourg: Language Policy Division. Council of Europe