

**VALIDEZ, CONFIABILIDAD Y AMENAZAS A LA VALIDEZ EN LAS
EVALUACIONES DE LA CLASE DE INGLÉS I VIRTUAL DEL
DEPARTAMENTO DE LENGUAS EXTRANJERAS DE LA UNIVERSIDAD
NACIONAL AUTÓNOMA DE HONDURAS**
**VALIDITY, RELIABILITY, AND THREATS TO VALIDITY IN VIRTUAL ENGLISH I CLASS
ASSESSMENTS IN THE FOREIGN LANGUAGES DEPARTMENT OF THE NATIONAL
AUTONOMOUS UNIVERSITY OF HONDURAS**

Bessy Valeska Mendoza Navas

Universidad Nacional Autónoma de Honduras, Honduras

(d.lenguasextranjer@unah.edu.hn) (<https://orcid.org/0000-0003-1898-256x>)

Leonel Armando Madrid Argenal

Universidad Nacional Autónoma de Honduras, Honduras

(d.lenguasextranjer@unah.edu.hn) (<https://orcid.org/0000-0002-8354-7858>)

Información del manuscrito:

Recibido/Received: 06/12/2023

Revisado/Reviewed: 12/02/2024

Aceptado/Accepted: 09/04/2024

RESUMEN

Palabras clave:

representaciones, validez,
fiabilidad, amenazas a la validez,
Inglés Virtual

El presente artículo corresponde a un estudio realizado en el marco de la tesis de maestría denominada: "Representaciones de estudiantes sobre las amenazas a la validez, y seguridad de las evaluaciones de la clase de Inglés I Virtual del Departamento de Lenguas Extranjeras (DLE) de la Universidad Nacional Autónoma de Honduras (UNAH)." con la finalidad de tener referencias que sirvan para eliminar amenazas y corregir brechas de seguridad a las evaluaciones de la clase de Inglés I virtual, y que al mismo tiempo sirva de antecedentes para desarrollar un sistema válido y lo más seguro posible de evaluación en clases virtuales de lengua del DLE de la UNAH. El estudio adopta una metodología mixta: los datos analizados fueron tomados de 2 instrumentos que se aplicaron a estudiantes de inglés I virtual del III periodo académico (PAC) 2022. Los instrumentos se llenaron en línea en forma de encuesta usando Microsoft Forms. El análisis cuantitativo se realizó con el software de IBM SPSS, y el análisis cuantitativo con Atlas/Ti. Los resultados de esta investigación han permitido validar la Hipótesis Nula: las evaluaciones de la clase de inglés I virtual no cuentan con la validez ni la seguridad que reflejen la adquisición de competencias lingüísticas alcanzadas por los estudiantes según el MCER correspondientes a 56 horas. Si bien la validez logra cumplir con 4 de 5 evidencias de validez sugeridas por Downing, 2003 y Messick, 1989; hay inferencias que ponen en riesgo la seguridad de las evaluaciones debido a la posibilidad abierta de trampas y plagio en las evaluaciones.

ABSTRACT

Keywords:

representations, validity,
reliability, threats to validity,
virtual English

This article corresponds to a study carried out within the framework of the Master's Thesis called: "Students' representations on threats to validity, and reliability of assessments of the Virtual English I class of the Department of Foreign Languages of the National Autonomous University of Honduras" with the purpose to have references that serve to eliminate threats and correct security gaps to assessments of the virtual English I class and at the same time give out the background to develop a valid and as secure as possible evaluation system in virtual language classes of the UNAH. The study adopts a mixed methodology, the instrument used to collect data were two questionnaires applied to students of virtual English I of the III PAC 2022. The instruments were filled out online in a survey form using Microsoft Forms. The quantitative analysis was performed with IBM Spss software the quantitative analysis with Atlas/Ti. The results of this research have allowed validation of the Null Hypothesis about the evaluations of the virtual English I class do not have the validity or security that reflects the acquisition of language skills achieved by students according to the CEFR corresponding to 56 hours. Although the validity manages to comply with 4 of 5 validity shreds of evidence suggested by (Downing, 2003; Messick, 1989), some inferences put the safety of the evaluations at risk due to the open possibility of cheating and plagiarism in the evaluations.

Introducción

Esta investigación tiene su génesis en las dudas que tienen los docentes de lenguas, específicamente los profesores que imparten la clase de inglés I virtual de la UNAH, acerca de la validez y seguridad de las evaluaciones en línea. Dudas acerca de no saber si efectivamente son los estudiantes quienes realizan las evaluaciones u otras personas, si los estudiantes copian o cometen plagio, o si modifican trabajos anteriores. Esto representaría un indicador falso sobre la validez de los resultados en materia de adquisición de competencias lingüísticas de forma virtual. Otras dudas surgen en torno a si se puede desarrollar y evaluar la producción y la interacción oral en línea.

La enseñanza, adquisición, y evaluación de competencias lingüísticas en entornos virtuales puede verse afectada por diversos factores como aspectos, formación tecnológica y lingüística previa, disposición de recursos tecnológicos, hábitos y disciplinas de estudio, tiempo disponible, carga y presiones académicas, formación de valores como la honestidad, la pertinencia de los programas y contenidos, la metodología y práctica apropiada de enseñanza virtual, diseño y aplicación de evaluaciones, realimentación, entre otros. Estas circunstancias hacen compleja la visualización de los efectos en el aprendizaje y sus resultados en el desarrollo de habilidades esperadas y de logros de competencia. Por eso es necesario definir, a partir de diagnósticos precisos y de investigaciones puntuales, los criterios, metodologías e instrumentos necesarios para una sólida y pertinente evaluación de los distintos componentes de un entorno virtual. El presente trabajo de investigación se justifica ante la necesidad de buscar soluciones a la validez y seguridad de las evaluaciones en la clase de inglés I virtual del departamento de Lenguas Extranjeras de la Universidad Nacional Autónoma de Honduras (UNAH) con el objetivo de reducir las amenazas a los procesos de evaluación en línea.

La evaluación genera múltiples reacciones en los estudiantes, no todas agradables. Solemos decir que las actividades organizacionales deben ser evaluadas, que para mejorar hay que evaluar, que sin evaluación es imposible saber exactamente en dónde estamos, entre otros enunciados, pero cuando se trata de que la evaluación sea personal la situación cambia, sobre todo cuando es sumativa de alto impacto. Tener noticia de que seremos evaluados, frecuentemente nos genera una sensación de incomodidad, de angustia o hasta miedo, y si podemos diferirla o exentarla en ocasiones, lo hacemos. (Sánchez, 2021).

Evaluar requiere una actitud reflexiva y madura, recursos para realizarla, personal con entrenamiento y experiencia en sus matices metodológicos y técnicos, tiempo para planearla, realizarla y analizarla, así como infraestructura, todo para documentar las diferentes etapas del proceso. Los procesos de evaluación efectivos requieren estructuras organizacionales participativas, no tan verticales o jerárquicas, que estén dispuestas a aceptar los resultados con entusiasmo y transparencia, para actuar en consecuencia y mejorar la estructura, procesos y resultados del sistema. Se requiere pensamiento sistémico y visión de largo plazo para que el proceso de evaluación se integre adecuadamente al sistema, y la participación activa de las personas que conforman los diferentes elementos del mismo. En resumen, la evaluación no es una tarea fácil ni sencilla, se requiere esfuerzo a nivel individual y colectivo, así como apoyo de los diversos niveles de la estructura organizacional. (Sánchez, 2021).

Revisión de la literatura

La introducción previa conduce a la revisión bibliográfica de los aspectos centrales abordados en este trabajo, con la finalidad de aportar argumentos de referencia que puedan avalar los resultados del estudio que se presenta.

Hay muchas definiciones del término “evaluación” en el área educativa, Miller la define como: “un término genérico que incluye un rango de procedimientos para adquirir información sobre el aprendizaje del estudiante y la formación de juicios de valor...” (Miller, 2012). Ello implica un proceso sistemático de acopio de información mediante la aplicación de diversos instrumentos, como son exámenes escritos u orales, para ser analizada con rigor metodológico y así fundamentar la toma de decisiones. La edición más reciente de los Estándares para Pruebas Educativas y Psicológicas de AERA-APA-NCME, define “evaluación” como: “método sistemático de obtención de información, usado para formular deducciones sobre las características de personas, objetos o programas; proceso sistemático para medir o evaluar las características o el desempeño de individuos, programas u otras entidades con la finalidad de hacer inferencias; en ocasiones se usa como sinónimo de prueba” (AERA, APA y NCME, 2014).

Independientemente de las definiciones técnicas que usemos de evaluación y sus conceptos cercanos, el profesorado que tiene interacciones con estudiantes debe internalizar la evaluación desde una visión más profunda, como sugirió en 1977 Rowntree, cuando dice que una persona obtiene e interpreta información de manera consciente sobre el conocimiento y la comprensión, habilidades y actitudes de la otra persona cuando ésta interactúa con otra persona directa o indirectamente.

En los últimos años ha adquirido fuerza el concepto de “evaluación del-para-como aprendizaje” (“assessment of-for-as learning” en inglés), que pretende modificar el énfasis que ha existido en la evaluación sumativa, exámenes y calificaciones, hacia un panorama más amplio e integrado que nos lleve a anclar todo el proceso de evaluación con el aprendizaje, la meta fundamental del proceso educativo (Ashford-Rowe et al, 2014; Bennett, 2015; Harapnuik, 2021; Maki, 2010; NFETLHE, 2017a). La enseñanza, el aprendizaje y la evaluación son conceptos y actividades que están ligados de manera inextricable, y el alineamiento entre estos elementos con la planeación, diseño e implementación curricular es indispensable y se convierte en un elemento clave para el éxito educativo.

- Evaluación DEL aprendizaje. De acuerdo con varios autores este tipo de evaluación es equivalente a la evaluación sumativa, para documentar que el aprendizaje ocurrió y el nivel de éste. Su naturaleza es evaluar actividades que ya ocurrieron, después o al final de un periodo de aprendizaje, y enfatiza los aspectos cuantitativos y numéricos, asociándose con las calificaciones o grados. Cuando esta evaluación tiene consecuencias importantes en el estudiante, se le denomina “evaluación de alto impacto”. En este tipo de evaluación el actor principal es el profesor o la organización que aplica la evaluación, quienes son los principales tomadores de decisiones, y el estudiantado es un participante pasivo que recibe o a quien se le aplica el examen o la prueba, en contraste con la evaluación para el aprendizaje.
- Evaluación PARA el aprendizaje. Como se comentó anteriormente, la principal meta de la evaluación debe ser mejorar el aprendizaje, no sólo medirlo, por lo que cuando hablamos de evaluación para el aprendizaje nos referimos a la evaluación tradicionalmente llamada formativa, ligada a la retroalimentación (Maki, 2010; Man Sze Lau, 2016; Martínez Rizo, 2009; Wiliam, 2011). Esta

evaluación ocurre durante todo el proceso de enseñanza y aprendizaje, es más longitudinal y representa un diálogo que ocurre entre docentes y estudiantes a lo largo de sus múltiples interacciones. Está enfocada en ayudar al estudiante, identificar sus áreas de oportunidad y logros, para orientarlo a progresar de una mejor manera en el proceso educativo, sin generarle estrés o desgaste, tratándolo como persona. Pretende moverse de una acción que se le hace *al* estudiante, a un proceso que se hace *con* el estudiante. Esta evaluación es inseparable de la enseñanza y apoya fuertemente el aprendizaje, si se lleva a cabo con profesionalismo y responsabilidad.

- Evaluación COMO aprendizaje. En este tipo de evaluación el estudiante es empoderado, tiene mayor responsabilidad en el proceso de aprendizaje y puede ser el tomador clave de decisiones. El estudiantado necesita adquirir habilidades para el uso de los conceptos básicos de evaluación en su desarrollo personal. El aprendizaje autodirigido a lo largo de la vida, el aprendizaje autónomo, el pensamiento crítico, entre otros, requieren evaluar datos e información sobre situaciones laborales y vivenciales, analizarlos, establecer juicios de valor, tomar decisiones sobre la problemática personal y profesional. Todo ello requiere autoevaluación y capacidad de tomar decisiones con base en evaluación de contextos y realidades complejas. Aunque el profesorado generalmente es quien tiene el poder jerárquico en el proceso educativo formal, la evaluación como aprendizaje mueve este *locus* de control externo a un control más intrínseco por parte del estudiantado. Sin embargo, el estudiante requiere apoyo de docentes y sus pares para ejercer cabalmente las habilidades mencionadas. La evaluación como aprendizaje ayuda a los estudiantes a aprender a aprender, fomenta la metacognición y el aprendizaje autorregulado.

Es muy importante que se revise lo que se está enseñando y lo que se espera que los estudiantes aprendan; la alineación de la enseñanza, el aprendizaje y la evaluación son esenciales en cualquier ámbito de la educación, pues de esta manera el docente garantiza que es lo que se está enseñando y aprendiendo sea efectivamente eso que se está evaluando. (Basabe et al. 2020).

Sería un esfuerzo perdido lo que se haga en la evaluación si no tenemos estos objetivos. La educación formal se guía por procesos claramente organizados por un currículo, plan de estudios, y programas de la asignatura. Estos planes y programas son la guía en el momento de nuestras evaluaciones y cerciorarnos de que efectivamente, eso que enseñamos y esperamos que hayan aprendido nuestros estudiantes, es lo que en efecto vamos a evaluar.

Así que es imprescindible conocer a cabalidad estos dos documentos para revisar detenidamente cada uno de sus elementos y la función o razón de ser de cada uno. En el caso del plan de estudios ubica a la asignatura y las conexiones que tiene con el resto de las asignaturas que lo integran. Con ello se puede orientar la evaluación de los objetivos, no únicamente de la signatura, sino también del conjunto del plan de estudios, contribuyendo así a la formación de los estudiantes. El plan de estudios es un sistema en el cual se derivan varios engranajes, y es necesario que todos funcionen adecuadamente para lograr los propósitos establecidos. Esto es precisamente a lo que nos referimos cuando hablamos de alineación de la enseñanza, el aprendizaje y evaluación.

Validez, confiabilidad y amenazas a la validez

A lo largo de la vida como docentes realizamos muchas evaluaciones para intentar conocer el nivel de conocimiento o desempeño de nuestros estudiantes. Este proceso implica la elaboración, aplicación e interpretación de exámenes de diferentes tipos: diagnósticos, formativos y sumativos. Independientemente de su finalidad, la meta de cualquier evaluación incluye la identificación del nivel de algún constructo, como competencia de comunicación escrita, competencia de comunicación oral o interacción en el caso de lenguas extranjeras.

Los resultados de las evaluaciones idealmente deben reflejar de una manera precisa y reproducible lo que se pretende evaluar, para poder interpretar de forma racional los resultados de la misma y estar capacitadas para realizar inferencias y tomar decisiones con fundamentos sólidos. Cuando se evalúa a los estudiantes sobre un tema particular, se desea identificar el proceso y resultados del aprendizaje que permitan inferir el nivel de desempeño en los constructos de interés. Después de aplicar las evaluaciones, obtenemos resultados en forma de puntuaciones que ayudan a tomar decisiones, que conllevan las siguientes interrogantes: ¿estamos evaluando exactamente lo que deseamos evaluar?, ¿qué implican los resultados con respecto al avance académico del estudiante?, si se trata de una evaluación sumativa, ¿cuál es la calificación mínima para aprobar el curso?, ¿qué tan reproducible es la medición?, entre muchas otras. La evaluación en educación es una disciplina cada vez más sofisticada y sustentada en investigación, que requiere incorporar conceptos académicos fundamentales para llevarse a cabo con profesionalismo y solidez metodológica (Instituto Nacional para la Evaluación de la Educación, 2017).

El pilar conceptual más importante de la evaluación en educación es la validez. Actualmente el concepto de validez ha evolucionado del tradicional “medir lo que se pretende medir”, a un modelo más amplio y profundo, en el que “se refiere al grado en que la evidencia y la teoría respaldan las interpretaciones de los puntajes de una prueba para usos propuestos de las pruebas” (AERA, APA y NCME, 2018). Es un conjunto de acciones que se ubican a lo largo del proceso evaluativo, para fundamentar la interpretación de los resultados y así generar inferencias. El análisis de validez, o validación, es el proceso mediante el que evaluamos las evidencias que se presentan para determinar cuál es el grado de validez (Cook y Hatala, 2016). Se puede realizar para los diferentes tipos de exámenes, diagnósticos, formativos y sumativos, aunque es particularmente relevante para las evaluaciones sumativas de alto impacto.

Tradicionalmente la validez en educación se clasificaba como “las 3 Cs”: validez de contenido, de criterio y de constructo (Cronbach y Meehl, 1955). En la definición actual esta distinción desapareció, ya que el modelo vigente propone diversas fuentes de evidencia que arrojan luz sobre distintos aspectos de la validez, no es que reflejen diferentes tipos de la misma. La validez es un concepto unitario, por lo que se considera que toda la validez es validez de constructo.

Posteriormente, a finales del siglo XX, un nuevo marco de referencia de validez fue propuesto y aceptado por las principales organizaciones de evaluación educativa y pruebas psicológicas (American Educational Research Association et al., 2018), incorporando el concepto holístico de validez de constructo. Este modelo establece que, para determinar el grado de validez de los usos e interpretaciones de los resultados de una evaluación, se deben proveer diversos elementos que lo demuestren (Downing, 2003). Este esquema propone los siguientes elementos como cinco fuentes de evidencia de validez (Downing, 2003; Messick, 1989):

- Evidencia basada en el contenido de la prueba.

- Evidencia basada en los procesos de respuesta.
- Evidencia basada en la estructura interna. La estructura interna presenta tres características básicas: dimensionalidad, funcionamiento diferencial y confiabilidad (Rios y Wells, 2014). Al diseñar la prueba, se debe determinar que dimensiones se desean evaluar sobre el constructo de interés, y esta información se describe en la tabla de especificaciones del examen.
- Evidencia basada en las relaciones con otras variables.
- Evidencia basada en las consecuencias de la prueba. Los resultados de las pruebas.

Validación

La validación es un proceso que se debe planear al mismo tiempo que se diseña la prueba, para asegurarse de contar con las fuentes de evidencia necesarias para obtener el mayor grado posible de validez de la interpretación de sus resultados. A continuación, se sugiere una manera de realizar este proceso.

1. Especificar los usos e interpretaciones de las puntuaciones

1a. Formular los usos e interpretaciones. Los usos y las interpretaciones de las puntuaciones que se obtienen en una prueba son conceptos diferentes y ambos se deben aclarar desde que se inicia el diseño de la prueba.

1b. Establecer las hipótesis. Las hipótesis son preguntas que nos podemos hacer acerca de la evaluación que se está elaborando. Deben probarse por medio de las fuentes de evidencia mencionadas.

2. Evaluar las fuentes de evidencia

2a. Crear un plan para probar las hipótesis. Con base en las hipótesis seleccionadas, se buscan las fuentes de evidencia y se reúne la información correspondiente.

2b. Evaluar la evidencia y formular un juicio. En este último paso se evalúan todas las evidencias en orden y se establece el grado de validez de la interpretación de las puntuaciones de la prueba evaluada. Este grado dependerá de la calidad de las evidencias presentadas y también de las evidencias más importantes, según la prueba.

Amenazas a la validez

Además de analizar las fuentes de evidencia de validez, se sugiere identificar elementos que puedan afectar el grado de validez de los resultados de la evaluación. Este paso es importante porque da fortaleza a las decisiones que se toman con base en los resultados de la prueba. A los elementos que reducen el grado de validez se les denomina amenazas a la validez; se les llama así porque interfieren con la correcta interpretación de las puntuaciones (Carrillo-Ávalos et al., 2020; Downing y Haladyna, 2004). Estas amenazas pueden estar presentes en cualquier tipo de evaluación. En general, se reconocen dos tipos de amenazas a la validez: subrepresentación del constructo (SC) y varianza irrelevante al constructo (VIC) (Downing, 2003; Messick, 1989).

La primera es la amenaza a la validez por subrepresentación del constructo. Esto se refiere a que hay una representación inapropiada de los dominios explorados en la evaluación del contenido a evaluar por los exámenes. Por ejemplo, cuando un examen tiene muy pocos reactivos o muy pocas preguntas, que no muestreen, que no exploren apropiadamente el área de conocimiento a revisar. Otro ejemplo es la distribución de reactivos que no sigan fielmente la tabla de especificaciones. De manera que algunas áreas acaban siendo sobre exploradas y otras infra exploradas. Incluso hay ocasiones que hay áreas que ni siquiera se exploran en algún examen. Esto, obviamente, afecta a la validez del uso del examen. Otro ejemplo son muchos ítems, muchas preguntas, que

exploren procesos cognoscitivos de bajo nivel, como la memoria o el reconocimiento de datos factuales, mientras que los objetivos de la enseñanza idealmente son de mayor nivel, como la aplicación o la solución de problemas. Otra amenaza a la validez, que ha adquirido cada vez más importancia, es el fenómeno de enseñar para la prueba. Esto significa que el profesor enfatiza demasiado en la clase lo que va a incluirse en el examen, distorsionando así el currículo, las metas educativas y en general, todo el proceso; esto ha llegado a ocurrir en grado tal que algunos profesores utilizan reactivos del examen en clase, para artificialmente aumentar las calificaciones de sus alumnos y así mejorar las evaluaciones de su grupo o incluso de su institución en este mundo de rendición de cuentas educativas.

El segundo gran tipo de amenazas a la validez es lo que llamamos varianza irrelevante al constructo. Esto se refiere a variables que, de manera sistemática, interfieren con la capacidad de interpretar los resultados de la evaluación de una manera significativa, y que causan, digamos, ruido en los datos de medición. Ejemplos de este tipo de amenaza a la validez son reactivos elaborados con deficiencias y que tienen fallas. El escribir buenos reactivos de preguntas de examen es un arte y una ciencia a la vez, y requiere entrenamiento y experiencia. No es tan fácil, como pensamos con frecuencia. Otro ejemplo son los problemas que ocurren con la seguridad del examen y con la fuga de información o el hacer trampa en el examen, el copiar, el utilizar esto que llamamos acordeones, de manera que el resultado del examen no refleja con fidelidad lo que realmente sabe la persona.

Esto invalida obviamente los resultados de los exámenes, con complejas implicaciones éticas y de uso de recursos, como es el repetir el examen, o realizar otra versión o tomar medidas represivas con los estudiantes. La mayor parte de los departamentos que ofrecen clases en línea en la UNAH tienen bancos de reactivos que no son muy grandes o no tienen un banco de reactivos, por lo que el sobreexponer las preguntas de los exámenes, se convierte en un gran problema operacional. Por otra parte, el crear una cultura punitiva o de castigo alrededor de las evaluaciones no es el mensaje que idealmente debemos dar los docentes a los estudiantes, por lo que estos aspectos deben tomarse en cuenta para plantear cómo responder cuando ocurran este tipo de irregularidades. Existe también algo que se llama la astucia para responder exámenes; esto ocurre cuando los estudiantes se preparan con estrategias para responder exámenes y pueden obtener puntajes que no reflejen necesariamente lo que saben, sobre todo en exámenes que no están bien hechos.

Método

La investigación que hemos llevado a cabo se inscribe dentro de los métodos de investigación en aprendizajes de lenguas como un estudio no experimental *ex post facto* de campo, de nivel MIXTO descriptivo, pero con una característica de tipo correlacional.

Participantes

Sobre una muestra autoseleccionada de 163 sujetos en corte transversal único, el estudio se ha llevado a cabo con estudiantes de 5 clases (Secciones: 0702, 0800, 1005, 1401, 1802) que cursaban inglés I general en el III periodo académico (PAC) 2022 en el Departamento de Lenguas Extranjeras de la Universidad Nacional Autónoma de Honduras.

Instrumento de investigación

Se aplicó un cuestionario mixto auto-administrado con ítems, tanto cerrados como semicerrados. La frecuencia de respuesta de cada ítem es presentada en forma tabular y gráficas. Finalmente, para las citas textuales se ha implementado la codificación abierta y la codificación selectiva.

Análisis de datos

Para el análisis e interpretación de los datos se escogieron diferentes categorías atendiendo los temas principales planteados en la investigación: la evaluación (E), seguridad en las evaluaciones (SE), el uso de herramientas tecnológicas (UHT), Mejoras (M) y nivel de satisfacción de la clase (NS). En cada categoría se discriminaron una serie de subcategorías respondiendo a indicadores aportados por los distintos informantes y vinculados directamente con los temas principales seleccionados de antemano, lo cual permitió hacer manejable el cúmulo de información recogida durante la investigación y presentar los resultados en función de los objetivos propuestos. En lo que respecta a la presentación de los resultados e interpretación de las preguntas abiertas utilizadas en el cuestionario que proporcionan información de carácter textual, opiniones, explicaciones, justificaciones, se implementó el análisis desde la perspectiva de las diferentes categorías entrándose en las respectivas subcategorías definidas.

Resultados

Con los hallazgos encontrados a lo largo de esta investigación pudimos validar nuestra hipótesis nula ***H₀***: Las evaluaciones de la clase de inglés I virtual del Departamento de Lenguas Extranjeras de la UNAH no cuentan con la validez ni la seguridad que reflejen la adquisición de competencias lingüísticas alcanzadas por los estudiantes según el MCER correspondientes a 56 horas.

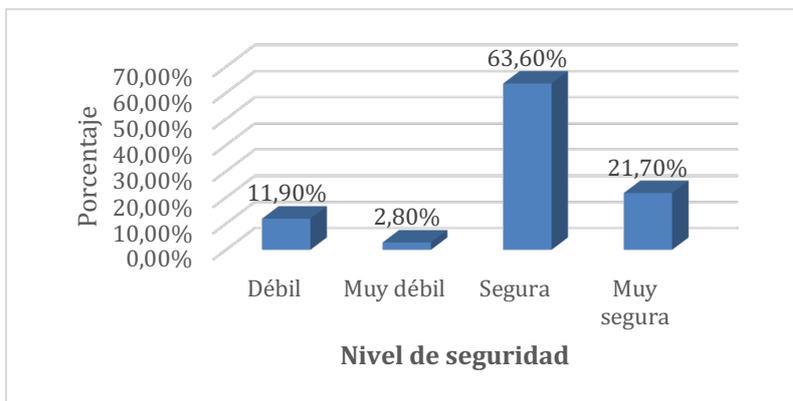
Los hallazgos revelaron que:

- Las evaluaciones totales del curso están en un rango de .854 de Alpha Cronbach.
- Las amenazas a la validez y a la seguridad de las evaluaciones de la clase de inglés I están en torno al 78,2%, lo cual demanda de actualización urgente de la clase.
- Si bien las evaluaciones muestran un porcentaje seguro en la elaboración, en el momento de ser respondidas, se presentan varias amenazas que necesitan ser resueltas de manera urgente.

Los análisis de inferencias que se realizaron a la validez de las evaluaciones, con relación a su variable de seguridad, dieron los siguientes resultados que muestran amenazas latentes a la validez de las evaluaciones:

- Un 14,7% de los encuestados manifestó que el nivel de seguridad para las evaluaciones y tareas era débil o muy débil.

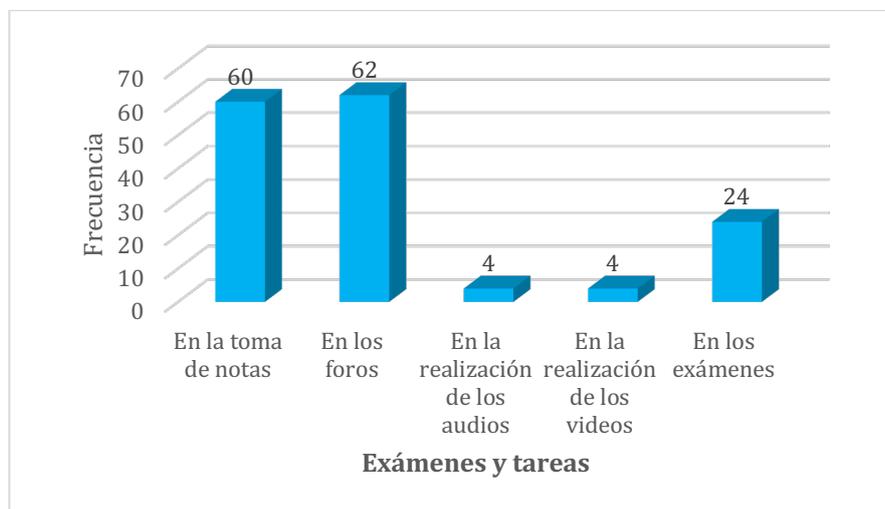
Figura 1
 Ítem C1. P12 nivel de seguridad



Las actividades en la que es más fácil hacer trampa o copiar según los encuestados, y que presentan amenazas a la validez son:

- Los foros (40%)
- La toma de notas (39%)
- Los exámenes (16%)
- Realización de audios (3%)
- Realización de videos (3%)

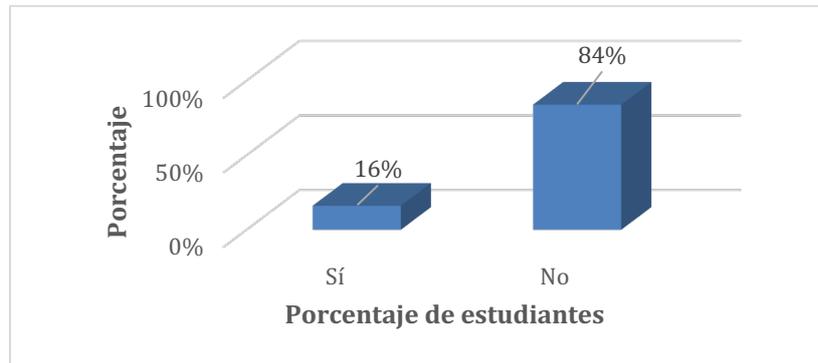
Figura 2
 Ítem C1. P16 examen o tarea más fácil de hacer trampa



Los encuestados manifestaron haber escuchado en un 16% que sus compañeros habían copiado o hecho trampa en las evaluaciones (C1.p18).

Figura 3

Ítem C1. P18 estudiantes de la clase de Inglés I copia o hacen trampa



Las repuestas a la pregunta C1.p18 muestran las formas en que los estudiantes hacen trampa en las que resaltan:

- El usar una tarea ya presentada en clases anteriores c1.p18-c,
- Pedir a otro(a) que le haga la tarea, familiar, o pagando a alguien bilingüe c1.p18-a,
- Tener abierto internet, libros, o apuntes al hacer el examen, cometer plagio por internet c1.p18-b,

Tabla 1

Ítem C1. P19 de qué forma hacen trampa en la clase de inglés I virtual

Formas de hacer trampa en las evaluaciones de Inglés I virtual, según los encuestados.			
Código	Forma	Frecuencia	%
c1.p18-a	Pidiendo a otro(a) que le haga la tarea, familiar, o pagando a alguien bilingüe.	45	25%
c1.p18-b	Teniendo abierto internet, libros, o apuntes al hacer el examen, Cometiendo plagio por internet	34	18.90%
c1.p18-c	Usando una tarea ya presentada en clases anteriores	55	30.60%
c1.p18-d	Trabajando con otro(a) en una tarea individual	16	8.90%
c1.p18-e	Usando herramientas digitales para modificar textos, cambiándole el nombre a las lecciones	9	5%
c1.p18-f	Usando herramientas digitales para modificar audios	2	1.10%
c1.p18-g	Usando herramientas digitales para modificar videos	2	1.10%
c1.p18-g	Usando herramientas de audio para fingir la voz	1	0.60%
c1.p18-i	Otras (sugeridas por los estudiantes)		
c1.p18-i-1	Compartiendo capturas de pantalla de exámenes, que otro le haga el examen o haciéndolo con otro, compartiendo revisiones con otros, usando teléfonos al momento de hacer los exámenes	11	6.10%
c1.p18-i-2	Usando traductor o buscando a alguien para que le traduzca	2	1.10%
c1.p18-i-3	Presentando tareas de clase anterior al repetirla	2	1.10%
c1.p18-i-4	Copiando respuestas de foros	1	0.60%
Total, de frecuencia y porcentajes		180	100%

Por lo anterior, se comprueba la Hipótesis Nula, si bien la validez logra cumplir con 4 de 5 evidencias de validez, hay inferencias que ponen en riesgo la seguridad de las evaluaciones debido a la posibilidad abierta de trampas y plagio en las evaluaciones.

Por otra parte, los resultados revelaron que a pesar de las amenazas a la validez en las evaluaciones, la clase de inglés I virtual ayudó a los estudiantes a:

- La producción escrita, la clase contiene 12 video lecciones de alrededor de 1 hora cada video (12hrs en total). Los estudiantes deben ver las lecciones, y tomar notas a partir de una guía de toma de notas para cada lección. Además, los estudiantes debían realizar tareas escritas en foros.
- Mejorar la comprensión oral y escrita al ver los videos. Las video-lecciones además de ayudar a mejorar la escritura, también ayudó a mejorar la comprensión oral y lectura.
- Realizar producciones orales a través de audio, video. La producción oral o *speaking* se desarrolló y se evaluó a través de la realización de audios a través de una herramienta simple (Vocaroo.com), y a través de la realización de

videos para asegurarse que la persona en el video es el/la estudiante quien realiza la producción oral.

- Desear aprender más de la lengua. El dato es relevante porque se sembró a través de la clase un deseo de aprender más de la lengua.

Otros aportes que realizó la clase fueron mejorar la lectura, entablar una conversación simple, ganar confianza al hablar, a lograr autonomía en el aprendizaje, desarrollar habilidades tecnológicas y ser más organizado en el tiempo de estudio. Un porcentaje menor al 1% (0,5%) expresó que no le había ayudado en nada la clase, y un 0,5% dijo que le habían ayudado en otros aspectos, pero no mencionó cuáles eran esos aspectos. A pesar de presentar amenazas a la validez y a la seguridad en las evaluaciones, la clase hizo aportes a la formación lingüística de los(as) estudiantes, y a desarrollar valores, y motivación para aprender más de la lengua.

Discusión y conclusiones

Esta investigación tenía como objetivo principal analizar las representaciones de estudiantes de Inglés I - III PAC 2022 del Departamento de Lenguas Extranjeras de la UNAH sobre las amenazas a la validez y a la seguridad de las evaluaciones en línea de la clase de Inglés I virtual. La investigación se realizó para tener datos que sirvan para eliminar amenazas y corregir brechas de seguridad a las evaluaciones de la clase de Inglés I virtual, y que al mismo tiempo sirva de referencia para desarrollar un sistema válido y lo más seguro posible de evaluación en clases virtuales de lengua del DLE de la UNAH. El investigador partió de la comprobación de las 2 hipótesis siguientes:

1. Hipótesis Nula ***H₀***: Las evaluaciones de la clase de inglés I virtual del DLE de la UNAH NO cuentan con la validez ni la seguridad que reflejen la adquisición de competencias lingüísticas adquiridas por los estudiantes según el MCER correspondientes a 56 horas.
2. Hipótesis Alternativa ***H_a***: Las evaluaciones de la clase de inglés I virtual del DLE de la UNAH SÍ cuentan con la validez y la seguridad que reflejen la adquisición de competencias lingüísticas adquiridas por los estudiantes según el MCER correspondientes a 56 horas.

Para comprobar las hipótesis, se siguió la forma de Hipótesis nula sugerida por Sheaham (2006), y cinco fuentes de evidencia de validez sugeridas por Downing, 2003; Messick y 1989:

1. Evidencia basada en el contenido de la prueba. (Las evidencias descansan en las propuestas de Sireci y Faulkner-Bond, 2014):

- a. Definición del dominio. Se analizó la descripción detallada de las áreas de contenido y las habilidades cognitivas que se deseaba evaluar del constructo definido en el currículo y de los resultados de la actividad de aprendizaje. Se encontró que las áreas de contenido están enmarcadas en los descriptores del MCER con las habilidades lingüísticas cognitivas propias de un nivel A1 del mismo Marco.

- b. Representación del dominio. Se analizó en las pruebas si las preguntas se habían establecido conforme a los objetivos o metas de aprendizaje y se encontró que sí lo estaban.

- c. Relevancia del dominio. Se encontró que los ítems eran importantes con respecto al aspecto del constructo que se está midiendo en la clase.

- d. Procedimientos apropiados de diseño de la prueba. Los ítems de las pruebas fueron probados antes de iniciar la clase en proyectos piloto. La revisión de ítems de las

pruebas se hace en cada período por expertos en contenido para asegurarse de su veracidad técnica. Ellos verifican que estén bien elaborados.

e. Credenciales de los creadores del examen, elaboradores de reactivos y expertos en contenido. Las evaluaciones fueron elaboradas por expertos en la enseñanza de lenguas y expertos en diseño de contenidos y de manejo de plataformas virtuales de la Dirección de Innovación Educativa (DIE).

2. Evidencia basada en los procesos de respuesta. Aunque en las evaluaciones de la clase de Inglés I, existen exámenes con preguntas de selección múltiple, las evaluaciones demandan producciones orales y escritas para verificar si el estudiante aplica los conocimientos adquiridos a la vida real (presentarse, hablar de sus actividades, de su familia, entre otros). la validación de la hoja de respuestas correctas, el control de calidad del reporte de los resultados, entre otros, como lo sugiere (Downing, 2003), fue realizada por expertos del DLE y de la DIE.

3. Evidencia basada en la estructura interna. Se analizaron tres características básicas conforme lo sugiere, Ríos y Wells, 2014:

a. Dimensionalidad. (Comprensión oral y escrita, producción oral y escrita, interacción).

b. Funcionamiento diferencial (Leenen,2014; Rios y Wells, 2014). Se analizaron 8,505 resultados de 45 exámenes aplicadas a 189 estudiantes de 5 diferentes secciones de inglés I- III PAC 2022. La misma prueba se aplicó tanto a hombres como a mujeres de diferentes edades.

c. Confiabilidad. Se realizó un *análisis de escala de fiabilidad de Alfa Cronbach con un resultado alto de .854*.

4. Evidencia basada en las relaciones con otras variables. No se pudo realizar otra prueba con estándares internacionales como IELTS, TOEFL, TOEIC, Cambridge, u otros, por cuestiones financieras.

5. Evidencia basada en las consecuencias de la prueba. Para este efecto se consultó a los encuestados sobre lo que más les había ayudado en la clase c1.p25 aportando evidencia de las consecuencias de las pruebas. Para futuros trabajos de investigación sobre la misma problemática, se recomienda, como lo hace Lane, 2014, realizar entrevistas, grupos focales, cuestionarios, para conocer cuáles son los componentes más importantes de los programas académicos y sus puntos de mayor impacto en el área del conocimiento de las lenguas.

Después del análisis, se concluye que la validez cumplió con 4 de las 5 evidencias, no se puede decir si cumple o no con la evidencia # 4 basada en las relaciones con otras variables ya que no se pudo analizar con otras variables porque no se cuenta con un examen similar que sea accesible económicamente a la totalidad de los encuestados.

A pesar que se comprobó que las evaluaciones si cumplen con la variable de validez de las evaluaciones según 4 de las 5 evidencias sugeridas por Downing, 2003 y Messick, 1989, también se debía analizar la variable de seguridad en las evaluaciones de forma que la validez contara con “un juicio valorativo holístico e integrador que requiere múltiples fuentes de evidencia para su interpretación”, y “que intente responder a la pregunta: ¿qué inferencias pueden hacerse sobre la persona basándose en los resultados del examen?” (Downing, 2003).

Los análisis de inferencias que se realizaron a la validez de las evaluaciones, con relación a su variable de seguridad, dieron los siguientes resultados que muestran amenazas latentes a la validez de las evaluaciones:

- Un 14,7% de los encuestados manifestó que el nivel de seguridad para las evaluaciones y tareas era débil o muy débil.

- Los foros, la toma de notas, fueron las actividades en las que es más fácil hacer trampa o copiar según los encuestados, y que presentan amenazas a la validez.
- Los encuestados manifestaron haber escuchado en un 16% que sus compañeros habían copiado o hecho trampa en las evaluaciones (C1.p18).
- Las repuestas a la pregunta C1.p19 muestran las formas en que los estudiantes hacen trampa, en las que sobresalen:
 - el usar una tarea ya presentada en clases anteriores c1.p18-c,
 - pedir a otro(a) que le haga la tarea, familiar, o pagando a alguien bilingüe c1.p18-a,
 - tener abierto internet, libros, o apuntes al hacer el examen, cometer plagio por internet c1.p18-b,
 - otras formas de hacer trampa se describen en la tabla descriptiva de la pregunta C1.p18 que resume las respuestas de los encuestados.

Por lo anterior, se comprueba la Hipótesis Nula, si bien la validez logra cumplir con 4 de 5 evidencias de validez, hay inferencias que ponen en riesgo la seguridad de las evaluaciones debido a la posibilidad abierta de trampas y plagio en las evaluaciones.

Recomendaciones

Para asegurar la validez y la seguridad en las evaluaciones, se recomienda:

1. Actualizar los programas, metodología, y evaluación de la clase de inglés I virtual.
2. Crear diferentes versiones de exámenes con un banco amplio de preguntas y respuestas de forma que se apliquen diferentes tipos de exámenes.
3. La revisión de las tareas en toma de notas debe ser de forma presencial para reducir la frecuencia de que se pasen y/o modifiquen los archivos digitales. Las tomas de notas deben quedarse con el profesor. En los foros debe enviarse la respuesta antes de ver los aportes de los/as compañeros. Que las tareas en audios se pasen a video para evitar que se editen. Si el profesor tiene dudas en la autenticidad de un trabajo que compare la voz y al/a la estudiante con el video de presentación que debe ser obligatorio para poder empezar a hacer las tareas. Que se configure la plataforma de manera que el estudiante no pueda avanzar si no tiene una nota en el foro de presentación.
4. Realizar los exámenes a la hora de la clase para evitar que se pasen las preguntas y dar la realimentación en un tiempo prudencial después para evitar que se visualicen las respuestas o dar una retroalimentación global vía Zoom de las correcciones, sin dejar las respuestas abiertas.
5. Debe crearse, socializarse, y aplicarse un Reglamento para sancionar el fraude o plagio apegado a las Normas Académicas y al Reglamento Estudiantil de la UNAH.
6. Acortar la duración de las video-lecciones, segmentarlas, para la toma de notas.
7. Realizar una sola evaluación presencial con valor de 60%, sea oral y/o escrita, y que el valor de las evaluaciones y tareas se convierta a un 40%. Esto hará que el estudiante se preocupe en llegar con más preparación lingüística al examen final.
8. Que las evaluaciones se hagan a la hora de la clase con cámara encendida, y con la supervisión del docente, si es en una plataforma especializada en

- exámenes mejor. Para ello habrá que incluir en el Plan Operativo Anual un presupuesto para licencias.
9. Crear un programa de formación de valores sobre la honestidad, ética, y responsabilidad que se incluyan en el curso con el fin de reducir o eliminar el fraude.
 10. Crear un plan de reuniones, asistencia, supervisión de exámenes, realimentación para estudiantes por parte de los asesores.
 11. Mantener reuniones constantes con el personal que imparte la clase para identificar, reportar, y corregir amenazas a la validez y a la seguridad de las evaluaciones, creación de recursos para reemplazar aquellos que necesiten ser actualizados.

Referencias

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, y Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. AERA. <https://www.testingstandards.net/open-access-files.html>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2018). *Estándares para pruebas educativas y psico lógicas*. American Educational Research Association.
- Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, 39(2), 205-222. <https://doi.org/10.1080/02602938.2013.819566>
- Basabe, L., Leal Falduti, B., & Tornese, D. (2020). *Diseño de exámenes con ítems de respuesta abierta*. Citep. Centro de Innovación en Tecnología y Pedagogía. http://citep.rec.uba.ar/wp-content/uploads/2020/05/AcaDocs_D09_Dise%C3%B1o-de-ex%C3%A1menes-escritos-con-%C3%ADtems-de-respuesta-abierta-1.pdf
- Bennett, R. E. (2015). The Changing Nature of Educational Assessment. *Review of Research in Education*, 39(1), 370-407. <https://doi.org/10.3102/0091732X14554179>
- Carrillo-Avalos, B. A., Sánchez-Mendiola, M., & Leenen, I. (2020). Amenazas a la validez en evaluación: implicaciones en educación médica. *Investigación en Educación Médica*, 9(34), 100-107. <https://doi.org/10.22201/facmed.20075057e.2020.34.221>
- Cook, D. A., & Hatala, R. (2016). Validation of educational assessments: a primer for simulation and beyond. *Advances in Simulation*, 1(1), 1-12. <https://doi.org/10.1186/s41077-016-0033-y>
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302. <https://doi.org/10.1037/h0040957>
- Downing, S. M. (2003). Validity: On the meaningful interpretation of assessment data. *Medical Education*, 37(9), 830-837. <https://doi.org/10.1046/j.1365-2923.2003.01594.x>
- Downing, S. M. & Haladyna, T. M. (2004). Validity threats: Overcoming interference with proposed interpretations of assessment data. *Medical Education*, 38(3), 327-333. <https://doi.org/10.1046/j.1365-2923.2004.01777.x>

- Harapnuik, D. K. (2021). Assessment OF/FOR/AS Learning. [Blog] It's About Learning. Creating Significant Learning Environments. https://www.harapnuik.org/?page_id=8900
- Instituto Nacional para la Evaluación de la Educación. (2017). Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación. *Diario Oficial de la Federación*. <https://www.inee.edu.mx/wp-content/uploads/2019/04/P1E104.pdf>
- Maki, P. L. (2010). *Assessing for Learning: Building a Sustainable Commitment across the Institution*. Stylus Publishing, LLC.
- Man Sze Lau, A. (2016). "Formative good, summative bad?" – A review of the dichotomy in assessment literature. *Journal of Further and Higher Education*, 40(4), 509-525. <https://doi.org/10.1080/0309877X.2014.984600>
- Martínez Rizo, F. (2009). Evaluación formativa en aula y evaluación a gran escala: hacia un Sistema más equilibrado. *Revista Electrónica de Investigación Educativa*, 11(2). <http://redie.uabc.mx/redie/article/view/231>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (pp. 13–103).
- Miller, M. D., Linn, R. L., & Gronlund, N. E. (2012). *Measurement and Assessment in Teaching* (11^a Ed.). Pearson.
- McKean M. & Aitken E.N. (2016) Educational Renovations: Nailing Down Terminology in Assessment. In: Scott S., Scott D., Webber C. (eds) *Leadership of Assessment, Inclusion, and Learning. The Enabling Power of Assessment*. Springer, Cham. https://doi.org/10.1007/978-3-319-23347-5_2
- National Forum for the Enhancement of Teaching and Learning in Higher Education (2017a). *Assessment OF/FOR/AS Learning*. <https://www.teachingandlearning.ie/our-priorities/student-success/assessment-of-for-as-learning/>
- Ríos, J. & Wells, C. (2014). Evidencia de validez basada en la estructura interna. *Psicothema*, 26(1), 108–116. <https://doi.org/10.7334/psicothema2013.260>
- Rowntree, D. (1977). *Assessing students: How shall we know them?* Kogan Page Ltd.
- Sánchez, M. & Martínez, A. (Eds.) (2022). *Evaluación y aprendizaje en educación universitaria: estrategias e instrumentos*. UNAM.
- Sánchez, M. et al., (2021). *Evaluación del y para el aprendizaje a distancia: Recomendaciones para docentes de educación media superior y superior*. CUAIEED, UNAM. <https://cuaieed.unam.mx/descargas/investigacion/evaluacion-del-y-para-el-aprendiza-je-V02.pdf>
- Sireci, S. & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26(1), 100–107. <https://doi.org/10.7334/psicothema2013.256>
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14, <https://doi.org/10.1016/j.stueduc.2011.03.001>.