

## **Conectando Diseño y Resultados: Perspectivas sobre un Examen de Competencia de Comprensión Lectora**

### **Bridging Design and Outcomes: Student Perspectives on a Reading Comprehension Proficiency Exam**

Marisol de la Caridad Patterson Peña <sup>1\*</sup>, <http://orcid.org/0000-0003-2372-3458>

Yoan Martínez Márquez<sup>1</sup>, <http://orcid.org/0000-0002-1950-567X>

[yoanm@uci.cu](mailto:yoanm@uci.cu)

Guillermo Manuel Negrín Ortiz <sup>1</sup>, <http://orcid.org/0000-0001-8637-9258>

[gmnegrin@uci.cu](mailto:gmnegrin@uci.cu)

Claudia Harsch <sup>2</sup>, <http://orcid.org/0000-0001-6864-1574>

[harsch@uni-bremen.de](mailto:harsch@uni-bremen.de)

<sup>1</sup>Universidad de las Ciencias Informáticas. Kilómetro 2 ½, Carretera San Antonio de los Baños, La Habana, Cuba.

<sup>2</sup> Universidad de Bremen, Alemania.

\* Autor para correspondencia. [marisol@uci.cu](mailto:marisol@uci.cu)

#### **RESUMEN**

El presente estudio, enmarcado en el proyecto “*Desarrollo de un sistema de certificación de dominio del inglés para la Educación Superior Cubana*”, tiene como objetivo analizar la retroalimentación recopilada a través de una encuesta, tras la aplicación de un instrumento de evaluación para medir la comprensión lectora a una muestra de estudiantes en la Universidad de las Ciencias Informáticas. Los datos recopilados se analizaron mediante estadísticas descriptivas, tabulación cruzada y técnicas de visualización para identificar tendencias y desajustes entre el contenido del examen y las expectativas de los examinados. Estas perspectivas pueden contribuir a ajustes continuos para la aplicación del examen de certificación en un piloto nacional a desarrollar en el 2026, destacando la participación de

los actores involucrados como clave para garantizar la calidad del examen. El estudio subraya la importancia del diseño participativo, demostrando cómo las perspectivas de los aprendices pueden mejorar su adaptación a contextos reales.

**Palabras clave:** análisis de datos, examen de certificación, encuestas post-examen, retroalimentación

### ***ABSTRACT***

The present study, framed within the project “*Desarrollo de un sistema de certificación de dominio del inglés para la Educación Superior Cubana*” (Development of an English Proficiency Certification System for Cuban Higher Education), is aimed at analyzing the feedback collected through a post-test survey administered after a proficiency reading test at the University of Informatics Sciences, with the goal of improving reading proficiency assessments in our local context. Data from this post-test survey sample were analyzed through descriptive statistics, cross-tabulation, and visualization techniques to identify trends and misalignments between test content and expectations. These insights can contribute to ongoing refinements for a national pilot study scheduled for April 2026, emphasizing stakeholder engagement as critical to ensuring exam quality. The study underscores the importance of participatory design in certification exams, demonstrating how learner perspectives can enhance alignment with real-world contexts.

**Keywords:** data analysis, proficiency exam, post exam survey, feedback

Recibido: 25/11/2025

Aceptado: 15/12/2025

## **INTRODUCCIÓN**

En una era de globalización, el dominio del inglés ha surgido como una habilidad crítica para la comunicación académica, profesional e intercultural. Marcos de referencias reconocidas internacionalmente, como el Marco Común Europeo de Referencia (MCER), subrayan la necesidad de alinear las evaluaciones lingüísticas con tareas comunicativas del

mundo real para garantizar su validez y relevancia. (Council of Europe, 2003; North, 2023). La investigación sobre las evaluaciones de competencia lingüística destaca el papel de las pruebas estandarizadas para diagnosticar fortalezas y debilidades en los sistemas educativos.

La certificación del dominio del inglés como lengua extranjera se ha erigido como un requisito académico y profesional crucial a nivel global. En este marco, el Ministerio de Educación Superior (MES) ha priorizado el desarrollo de un sistema estandarizado de certificación del inglés para alinear los resultados educativos nacionales con los estándares internacionales. Esta iniciativa responde a la doble demanda de preparar a los estudiantes para las oportunidades globales y preservar los requisitos académicos locales. El proyecto “Desarrollo de un sistema de certificación de dominio del inglés para la Educación Superior Cubana” ejemplifica este esfuerzo, adoptando una metodología inspirada en el Ciclo de Desarrollo de Tests de Lenguas del Consejo de Europa, definido en sus manuales fundamentales, los cuales incluyen encuestas de retroalimentación post-pilotaje, como parte del proceso de validación. (Council of Europe, 2011).

La primera fase del proyecto antes mencionado (2017-2022) implicó la realización del primer piloto nacional de un examen de competencia en el año 2022. Este fue diseñado con las tareas desarrolladas por los miembros del proyecto y la realización de encuestas post-examen para evaluar las percepciones de los examinados sobre el diseño, las dificultades percibidas y la relevancia de las tareas. Como parte de las fases del proyecto, se realizó el procesamiento estadístico de las tareas pilotadas para su posterior análisis. Sobre la base de estas percepciones, la segunda fase del proyecto se centra en revisar las tareas pilotadas para su refinamiento y así dar paso a la ejecución de un segundo piloto nacional de los instrumentos en desarrollo en mayo de 2026. (Martínez y otros, 2025).

El diseño y perfeccionamiento de instrumentos evaluativos de alto impacto, como los exámenes de certificación, es un proceso complejo que enfrenta desafíos inherentes. La claridad de las instrucciones, la adecuación del tiempo, la percepción de dificultad, la relevancia de las tareas entre otros aspectos, pueden afectar el desempeño, comprometiendo la validez aparente (*face validity*) y potencialmente, la validez de constructo del examen.

Identificar y corregir estas discrepancias es esencial, pero requiere mecanismos ágiles y centrados en la experiencia del evaluado.

Es por ello que la retroalimentación recibida a través de encuestas post-examen también constituye una herramienta que permite recoger las percepciones, dificultades y sugerencias de los usuarios directos: los estudiantes examinados, resaltando de esta manera la transformación de datos subjetivos en evidencia objetiva para la futura toma de decisiones. No se trata solo de ver la opinión de los estudiantes sino poder utilizar el análisis estadístico como base empírica para guiar un rediseño efectivo.

La fase de estudio piloto del proyecto cubano “Desarrollo de un sistema de certificación de dominio del inglés para la Educación Superior Cubana” se alinea con las mejores prácticas internacionales para la evaluación lingüística. Los estudios pilotos (*pre-trials*) a pequeña escala, como se explica en el Manual para el Desarrollo de Pruebas Estandarizadas de Dominio del inglés en la Educación Superior Cubana, Versión Preliminar (Collada et al., 2023), son esenciales para recabar información cualitativa sobre el diseño de las tareas, asegurando así su refinamiento.

Los procedimientos generales para estos ensayos preliminares a pequeña escala delimitan los pasos a seguir en el proceso permitiendo a los investigadores recopilar datos empíricos sobre la validez de las tareas con el fin de mejorar su calidad antes de considerarlas para su pilotaje o implementación a nivel nacional. Uno de estos pasos considera el uso de encuestas post-examen. La realización de este tipo de encuestas es una metodología crítica para recabar retroalimentación y mejorar la alineación de estos exámenes con las necesidades comunicativas reales de los examinados.

La mayoría de los expertos en el campo enfatizan el papel de las encuestas post-examen en la mejora de la autenticidad de las tareas, la equidad y la participación de los aprendices, fortaleciendo así el refinamiento iterativo de los exámenes de alto impacto (Bachman & Palmer, 2010; Hawkey, 2006).

A medida que el proyecto de evaluación avanza hacia la implementación nacional de su examen de competencia en 2026, este trabajo enfatiza la necesidad de integrar las perspectivas de las partes interesadas en el diseño de la evaluación proponiendo un marco participativo donde la retroalimentación de los examinados pueda informar nuevos

refinamientos iterativos. Este enfoque no solo mejora la autenticidad de las tareas, sino que también asegura que el impacto del examen en las prácticas de enseñanza y aprendizaje sea intencional y constructivo, un componente crítico de la validez (Bachman & Palmer, 2010; Gokturk & Dina, 2022; Rahman, 2017).

El objetivo central del estudio es, por lo tanto, desarrollar el análisis estadístico de la encuesta post examen aplicada tras la aplicación de exámenes de comprensión lectora como base empírica para guiar el rediseño efectivo de los exámenes de certificación de inglés del MES. El estudio busca analizar la retroalimentación recopilada y demostrar cómo la interpretación estadística rigurosa de la misma permitiría entre otros aspectos ayudar al equipo de desarrolladores a identificar cualquier desajuste entre el diseño de la prueba y las experiencias de los examinados, en el proceso de mejora del diseño evaluativo de las tareas de comprensión lectora en nuestro contexto local.

## **METODOLOGÍA Y MÉTODOS**

Para la realización del estudio y en consonancia con las diferentes fases del proyecto antes mencionado, en marzo de 2025 se llevó a cabo un ensayo (pre-trial) adicional a pequeña escala para obtener información complementaria sobre la validez del constructo de la tarea de lectura “R0004” pilotada durante el primer piloto nacional en 2022. Esta tarea había sido previamente revisada y refinada por parte de los miembros del proyecto de la Universidad de Ciencias Informáticas (UCI), teniendo en cuenta la retroalimentación obtenida a partir del procesamiento estadístico y las ideas discutidas en los talleres dirigidos por la mentora del proyecto, la Dra. Claudia Harsch.

La calificación de la tarea siguió las pautas estandarizadas elaboradas por los miembros del proyecto, publicadas en el Manual para el Desarrollo de Pruebas Estandarizadas de Dominio del inglés en la Educación Superior Cubana (Collada, I., et al, 2023). Los procedimientos se alinearon con los registrados en la versión preliminar de este manual.

Se procedió entonces a analizar el diseño y los resultados del cuestionario post-examen utilizado en el pilotaje de la primera fase del proyecto y publicado en la página 55 de la versión preliminar del Manual (Collada, et.al, 2023)

Para adaptar y refinar este cuestionario post-examen se tuvieron en cuenta los puntos claves sobre pilotaje y encuestas post-examen esbozados en los manuales del MCER (Consejo de Europa, 2011), así como la encuesta post-examen publicada en el informe técnico proporcionado a los investigadores del Instituto para el Progreso Educativo de la Universidad Humboldt (Rupp, Vock, Harsch y Köller, 2010).

La encuesta presentó un diseño de métodos mixtos al combinar preguntas con respuestas tipo escala Likert (preguntas 3 a 9), junto con respuestas numéricas de preguntas abiertas (preguntas 1 y 2) para triangular tendencias cuantitativas. Finalmente, las preguntas abiertas (preguntas 10 y 11) solicitaron percepciones cualitativas.

Dado que el estudio trata de comprender las percepciones de los examinados, se tuvo en cuenta en el análisis cómo las estadísticas descriptivas (incluyendo medias, moda, mediana y desviación estándar) proporcionaron información sobre las experiencias de los estudiantes. La moda es importante porque muestra las respuestas más comunes, indicando así opiniones prevalentes; la mediana da la tendencia central, que es útil para datos ordinales como las escalas Likert; mientras que la desviación estándar muestra la variabilidad de las respuestas. Estos análisis estadísticos pueden revelar si el examen se alinea con los antecedentes e intereses de los examinados, lo cual es clave para la equidad en la evaluación.

La metodología seguida tuvo en cuenta las mejores prácticas para encuestas post-examen efectivas. La encuesta se administró después del examen, mientras las experiencias estaban frescas, operacionalizando así el marco de validez sugerido por (Weir, 2005). Además, siguiendo las mejores prácticas de Anderson y Morgan (Anderson & Morgan, 2008), la misma se mantuvo por debajo de los 10 minutos para evitar o reducir la fatiga.

Para la realización de la encuesta se seleccionó una muestra de estudiantes de la propia universidad (UCI), administrándose de forma anónima a 206 estudiantes. La muestra por conveniencia fue estratificada por especialidad (Informática: 170; Ciberseguridad: 22; Bioinformática: 14) para mejorar la representatividad. Para minimizar el sesgo, la participación fue voluntaria y los objetivos se explicaron tanto por escrito como oralmente.

## **RESULTADOS Y DISCUSIÓN**

Las Preguntas 1 y 2 proporcionaron datos sobre la experiencia previa de los examinados. En la **Pregunta 1** *¿Es su primera vez haciendo este tipo de examen? De no ser la primera, ¿cuántas veces se ha presentado?* el mayor porcentaje de los examinados por primera vez, sugiere que el examen fue percibido como novedoso o administrado con poca frecuencia.

Entre aquellos que habían tomado el examen antes (156), la mayoría (72%) reportó uno o dos intentos, mientras que solo una minoría del 4% había realizado cuatro o cinco intentos previos. Un notable 24% de este subgrupo omitió las respuestas, potencialmente debido a opciones de respuesta ambiguas.

La **Pregunta 2** *De acuerdo a su percepción, ¿qué nivel cree usted tiene en la habilidad de comprensión de lectura?* medía la percepción de los examinados sobre su propio nivel de comprensión. De acuerdo a los resultados, A2 fue el nivel más auto-reportado (40.29%), seguido del nivel B1 (24.76%). Solo el 4.85% de los encuestados se identificó como avanzado (B2 o superior). Esto podría indicar que los estudiantes aún no se perciben a sí mismos cumpliendo con las expectativas a nivel universitario. Sobresale que un 19.42% se calificó por debajo de A2, con un 3.4% (7 encuestados) afirmando tener un nivel inferior a A1. Esto puede reflejar una conciencia limitada de los niveles del MCER o un desajuste entre los objetivos del examen y los resultados académicos previos de los examinados en cursos ya completados o en las pruebas de colocación.

Las preguntas de la 3 a la 6 se centraron en explorar cuatro dimensiones críticas de la experiencia del examinado: (1) la relación entre la familiaridad y el interés, explorando si los estudiantes expresaban un mayor compromiso con los temas que reconocían; (2) la correlación entre la familiaridad y la dificultad percibida; (3) la relación entre el interés y la dificultad, examinando si los temas considerados menos interesantes también se juzgaban como más difíciles; y (4) la satisfacción general, que evaluó las percepciones de los estudiantes sobre el diseño del examen, incluyendo la claridad de las instrucciones, la asignación de tiempo y la organización de las tareas.

La **Pregunta 3** *¿Qué nivel de familiarización tiene usted con los temas de los textos del examen?* consideró la familiaridad de los examinados con los temas del examen. El diseño del examen incorpora textos tanto cotidianos como más técnicos, para reflejar situaciones reales de lectura en entornos académicos y profesionales. Para el Ítem 1 "Anuncios", el

55.98% de los encuestados reportó baja familiaridad ("no familiarizado" o "poco familiarizado").

Para el ítem 2 "Comidas y Bebidas", el 59.70% expresó alta familiaridad ("bastante familiarizado" o "muy familiarizado"). Este segundo ítem, aunque familiar, recibió críticas de algunos encuestados quienes los consideraron "no interesante" o "poco interesante". No obstante este porcentaje fue muy bajo.

Para el ítem 3 "Detección de enfermedades usando Inteligencia Artificial", el 82.52 % de los encuestados reportó baja familiaridad "no familiarizado" (75 encuestados) o "poco familiarizado" (95 encuestados), marcándolo como el tema menos familiar. Un 36,40% de los participantes eligió la categoría de respuesta "No familiarizado", lo que podría sugerir irrelevancia para sus experiencias. Este tema, sin embargo, se alinea con las especificaciones de ciencia/salud no especializados que aparecen en la página 9 del manual antes citado (Collada, et al, 2023). Cabe señalar que, a pesar de la baja familiaridad, los encuestados lo encontraron muy interesante (ver Pregunta 4).

La mayor familiaridad (media  $\approx 2.78$ ) para el ítem 2 "Comidas y bebidas" en comparación con el ítem 3 "Detección de enfermedades usando IA" con una media  $\approx 1.85$  refleja relevancia cultural. Esto enfatiza la necesidad de contenido contextualizado localmente en los exámenes de competencia.

La **Pregunta 4** *¿Cuán interesante encontró usted los textos del examen?* se centró en el interés de los examinados por los textos. Un gran número de participantes calificó los tres temas como "bastante o muy interesantes", validando así su uso. Al analizar esta pregunta, es importante destacar que el ítem 3 "Detección de enfermedades usando IA" tuvo el mayor interés (84.95%) a pesar de la baja familiaridad con este tema ya vista en sus respuestas a la pregunta 3. Esto sugiere que la familiaridad no es el único impulsor del interés. El llamativo contraste en este tema en particular podría provenir de la especialización académica o profesional de los participantes, o puede reflejar la prominencia actual de la Inteligencia Artificial como tema de tendencia. El alto interés (media  $\approx 3.31$ ) a pesar de la baja familiaridad también podría sugerir que los examinados valoran temas novedosos, pero podrían necesitar andamiaje para interactuar con ellos. Los datos también sugieren que los exámenes localizados deben equilibrar familiaridad y desafío.

Al comparar las dos primeras opciones: "No interesante" o "poco interesante" en los tres temas, el ítem 1 "Anuncios" sobresale con el mayor número de estudiantes seleccionando estas dos opciones (87). Aunque el porcentaje es bajo, esto se alinea con algunos de los comentarios compartidos en las preguntas abiertas al final de la encuesta, donde también enfatizaron algunas valoraciones críticas respecto a este tema en específico.

Las respuestas a la **Pregunta 5** *¿Cuán difíciles encontró usted los textos del examen?* sirvieron para evaluar el nivel de dificultad percibido por los examinados. En general, la mayoría de los estudiantes encuestados categorizó los tres temas como "Fácil/No muy difícil": 56.31%, 53.40% y 50.49%, respectivamente, lo que puede indicar una tendencia positiva en la accesibilidad percibida.

Al analizar las relaciones clave en esta pregunta, es notable que el tema del ítem 3, "Detección de Enfermedades Usando IA", sea percibido como el más desafiante. El 34.31% de los encuestados lo calificó como "bastante difícil", mientras que el 9.80% lo describió como "muy difícil". Una alta media de dificultad (2.49) y una desviación estándar ( $DE \approx 0.81$ ) sugieren que este texto fue desafiante y polarizante para algunos de ellos.

Comparando la distribución entre la dificultad percibida y la familiaridad para este ítem 3, un 36.7 reportó no estar familiarizados, mientras que 70 estudiantes (34.31%) lo encontraron "bastante difícil", sugiriendo una relación potencial entre la baja familiaridad y una mayor dificultad percibida.

A pesar de tener la mayor familiaridad entre los tres ítems, el 2 "Comidas y Bebidas" mostró una dificultad moderada; un 53.40% seleccionó "no muy difícil", lo que sugiere que la complejidad del contenido influye en el desafío percibido independientemente de la familiaridad.

La **Pregunta 6** *¿Qué nivel de familiarización tiene usted con los métodos de evaluación usados en el examen?* midió la familiarización de los examinados con los formatos de evaluación utilizados. El nivel de familiaridad es bastante similar para ambos formatos: "Ejercicios de emparejamiento o enlazar" (54.36%) y "Ejercicios de opción múltiple" (60.19%), lo que sugiere que los estudiantes se sienten cómodos con ambos formatos estándar. Al comparar ambos tipos de ejercicios, el 44.17% de los examinados reportó baja familiaridad con el formato de ejercicios de emparejamiento; esto podría sugerir que se

sienten más cómodos con el formato de opción múltiple, ya que podrían estar más acostumbrados a hacer este tipo de ejercicios en clase o en otros exámenes.

A través de las preguntas 7-9, se analizaron factores logísticos como la longitud y el tiempo del examen, que afectan tanto la validez como la experiencia del estudiante.

La **Pregunta 7** *¿Cuán adecuado encontró usted lo siguiente: 1.1 tamaño de los textos, 1.2 tamaño del examen en su totalidad. 1.3 cantidad de tiempo para responder?* midió la percepción de los examinados sobre la idoneidad o adecuación del diseño del examen. El mayor porcentaje de participantes (85.78%) consideró que la longitud de los textos era adecuada. Una prueba de chi-cuadrado de bondad de ajustes muestra que el valor de la significación es  $p < 0.0001$ , lo cual indica que la opción seleccionada es estadísticamente muy significativa y no puede atribuirse al azar; esto por lo tanto respalda el diseño del examen. La longitud total del examen fue percibida como adecuada por un 73% de los encuestados. Sin embargo, no debe pasarse por alto que un 22.55% eligió las categorías "un poco/ o muy largo". Por lo tanto, esta es considerada el área ligeramente más crítica en comparación con el 12.25% que seleccionó la longitud de los textos individuales como "un poco/o muy largo". Este criterio también fue expresado por algunos encuestados en los comentarios proporcionados en las dos preguntas abiertas de la encuesta.

Los estudiantes de baja competencia tenían 1.7 veces más probabilidades (31,9% vs 18.9%) de percibir el examen en las categorías antes mencionadas. Sin embargo, esta diferencia no es estadísticamente significativa pues no existe una correlación significativa entre aquellos que se auto-evaluaron con niveles de competencia más bajos y los que percibieron el examen como demasiado largo.

En cuanto al tiempo disponible para responder, un 75,49% lo consideró mayormente adecuado, con solo unos pocos encuestados (41) que seleccionaron las dos primeras opciones: "muy corto" o "un poco corto".

En la **Pregunta 8** *¿Cuán satisfecho estuvo usted con lo siguiente?*, se destaca como un punto fuerte la alta satisfacción de los examinados con la organización de las secciones del examen (91,26%), lo que valida la estructura del examen. En cuanto a sus instrucciones, el 88,34% de los encuestados se sintió satisfecho, indicando que las instrucciones eran claras.

Este aspecto positivo también se enfatiza en las preguntas abiertas incluidas al final de la encuesta. En general, los examinados mostraron satisfacción con la logística del examen.

Aunque un alto porcentaje aprobó la organización del examen, algunos ajustes menores podrían mejorar aún más la accesibilidad para resolver los problemas mencionados por algunos de los estudiantes en su respuesta a las preguntas abiertas. Por ejemplo, el 3% de los encuestados criticó el tamaño de la fuente, señalando que era "demasiado pequeño" y otro 3% solicitó que los textos y las preguntas aparecieran en la misma página.

La **Pregunta 9** *En su opinión, ¿cómo siente que el examen midió su habilidad de comprensión de lectura en inglés?* proporcionó datos sobre la percepción de los examinados acerca de su efectividad. Las cuatro categorías de la escala ordinal de esta pregunta fueron codificadas numéricamente para calcular las medidas de tendencia central y dispersión. Es así que, al calcular las medidas estadísticas, la moda tuvo un valor de 3 = "bien" (con 130 respuestas, 63.1%) lo que indica que la opinión predominante entre los encuestados es positiva. La media  $\approx 3.19$ , sugiere que la mayoría de los examinados sintió que el examen medía efectivamente sus habilidades "bien" o "muy bien", lo cual indica una evaluación general favorable. El hecho de que la media sea mayor a 3 muestra que hay más respuestas en el extremo positivo. La Desviación Estándar ( $DE \approx 0.615$ ) es relativamente baja, indicando que las respuestas están concentradas alrededor de la media. Hay poca dispersión en las opiniones y la mayoría coinciden en una evaluación positiva. Solo el 8.26% consideró midió la habilidad "pobremente", cifra no significativa para el análisis.

Para el análisis de las preguntas abiertas 10 y 11, relacionadas con las percepciones cualitativas, las respuestas se agruparon en comentarios positivos, críticas y sugerencias. Los comentarios de los participantes se agruparon en categorías lógicas para resaltar temas recurrentes e ideas accionables.

Las respuestas a la **Pregunta 10** *¿Cree usted que haya maneras en que podamos mejorar el examen?* reflejan comentarios individuales, no encuestados únicos, ya que la mayoría de los examinados proporcionaron múltiples sugerencias. Solo cinco estudiantes simplemente respondieron con la palabra "no", sin añadir más comentarios; tres de los encuestados declararon explícitamente que no sabrían qué decir.

Es notable que, aunque algunas respuestas incluyeron evaluaciones críticas, un alto porcentaje de encuestados reconoció en sus comentarios el trabajo de los diseñadores del examen, felicitándolos y agradeciéndoles. Algunos incluso señalaron haber realizado ediciones anteriores y expresaron una mayor satisfacción con esta versión.

La retroalimentación positiva se centró principalmente en que el examen era adecuado y justo, con un énfasis adicional en su formato bien estructurado. Las críticas clave, aunque menos frecuentes, comparadas con la abrumadora retroalimentación positiva, se centraron en los siguientes temas: longitud del examen y limitaciones de tiempo, 18 menciones (8.2%); temas poco comunes o falta de variedad, 17 menciones (7.7%); dificultad con el vocabulario, 6 menciones (2.7%) y formatos de ejercicios, 6 menciones (2.7%).

Un porcentaje muy bajo de los encuestados (30 participantes) optó por no responder esta pregunta final de la encuesta, lo que podría significar que no tenían retroalimentación que ofrecer y por lo tanto optaron por no expresar nada.

En cuanto a las respuestas a la **Pregunta 11** *¿Tiene algún otro comentario que le gustaría compartir sobre el examen?* 56 examinados (30,9%) no tenían más comentarios. No obstante, algunos de estos encuestados expresaron la razón de su respuesta.

Las respuestas a esta pregunta reflejan comentarios individuales, y no encuestados únicos al igual que en la pregunta anterior. La retroalimentación positiva destacó cuatro temas principales, ponderados por la frecuencia de los comentarios. La porción más grande (40%) elogió la estructura, equidad y claridad del examen, con encuestados destacando frases como "muy bien hecho y estructurados", "está bien tal como está" e "interactiva y satisfactoria". Otro 30% valoró algunos puntos específicos sobre el formato, como la dinámica de los anuncios, la evaluación interactiva y los diversos formatos de preguntas, citando ejemplos como "interesantes los diversos formatos de pregunta", elogios por la inclusión de elementos visuales, "tiene ilustraciones" y "textos adecuados". Aproximadamente el 20% enfatizó la idoneidad del examen para estudiantes universitarios en niveles A2/B1, señalando que era "adecuado para estudiantes universitarios" y su capacidad para "determinar el nivel de inglés de A2 o B1". El 10% restante expresó satisfacción general sin detalles específicos, con comentarios como "no necesita cambios".

Las respuestas a esta pregunta abierta concernientes a críticas y sugerencias revelaron seis temas clave. Primero, el contenido y los temas atrajo las críticas más frecuentes, con participantes citando problemas como "Temas poco interesantes", "el tema de comidas y bebidas resulta complejo" y "textos científicos" como irrelevantes o demasiado técnicos.

En segundo lugar, se destacaron problemas de formato y diseño, incluyendo quejas como "letra pequeña", "examen visualmente cargado" y "textos separados de las preguntas".

En tercer lugar, surgieron preocupaciones sobre la longitud del examen y el tiempo, aunque solo 18 menciones señalaron que el examen se sentía "un poco largo" o solicitaban "textos menos extensos", así como "dar más tiempo". Esta retroalimentación crítica con pocas menciones, algunas de examinados de baja competencia, no exige realmente a los autores actuar sobre ellas.

En cuarto lugar, se criticó la claridad de las preguntas por opciones de "respuestas ambiguas" y redacción excesivamente complicada "preguntas enredadas". Estos criterios también tuvieron muy pocas menciones.

En quinto lugar, el vocabulario y la dificultad generaron quejas sobre "palabras rebuscadas" y una percepción de desajuste con el nivel A2/B1 "dificultad elevada para A2/B1". Finalmente, otras sugerencias incluyeron solicitudes de nuevos formatos de pregunta; por ejemplo, ejercicios de llenar espacios en blanco con verbos y la solicitud de exámenes más frecuentes o clases preparatorias para alinear mejor las expectativas.

## CONCLUSIONES

El análisis de los datos de la encuesta tuvo como objetivo descubrir cómo los elementos del diseño de las tareas influyeron en las experiencias de los aprendices para identificar áreas potenciales de refinamiento. En general, los datos reflejaron una percepción predominantemente positiva del diseño y validez del constructo. Un alto porcentaje de los encuestados estuvo de acuerdo en que el examen medía efectivamente sus habilidades de lectura, validando así su propósito: evaluar la competencia de comprensión de lectura en inglés, alineada con los niveles A2-B1 del MCER. El análisis de los datos de la encuesta también mostró una alta satisfacción con la organización del examen y la claridad de las instrucciones, destacando por tanto la calidad de su diseño como un éxito del piloto.

Al basar los hallazgos en estadísticas, se puede transformar la retroalimentación subjetiva en ideas accionables, haciendo de este artículo un recurso valioso para mejorar las evaluaciones de comprensión de lectura en el contexto local. Los resultados subrayan el valor de la evaluación basada en datos para diagnosticar fortalezas, abordar debilidades y alinear los exámenes con los objetivos pedagógicos y las experiencias de los examinados. Finalmente, incorporar este tipo de análisis de encuestas de retroalimentación a otros exámenes de competencia no solo proporcionará evidencia empírica de la validez de las tareas diseñadas, sino que también equipará al equipo del proyecto para refinar sistemáticamente los exámenes basándose en las percepciones de los examinados, asegurando así que sigan siendo sólidos, equitativos y educativamente impactantes.

## REFERENCIAS BIBLIOGRÁFICAS

- Anderson, P., & Morgan, G. (2008). *Developing tests and questionnaires for a national assessment of educational achievement* (Vol. 2). World Bank Publications. Available at: <https://openknowledge.worldbank.org/handle/10986/6524>
- Bachman, Lyle F., and Adrian S. Palmer. 2010. *Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World*. Oxford: Oxford University Press Gokturk Saglam, Asli.
- Collada Peña, I., et al (2023) *Handbook for Standardised Proficiency Test Development in Cuban Higher Education. Preliminary Version*. Cuban Language Assessment Network (CLAN) Ediciones Futuro.
- Council of Europe. (2003). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR): A manual*. Strasbourg: Council of Europe Publishing. ISBN 978-92-871-7159-6. Disponible en: <https://www.coe.int>
- Council of Europe (2011). *Manual for language test development and examining: For use with the CEFR*. Strasbourg: Council of Europe Publishing. ISBN 978-92-871-7590-7. Disponible en: <https://www.coe.int>

- Gokturk, A. & Dina, T. (2022). Evaluating Perceptions towards the Consequential Validity of Integrated Language Proficiency Assessment. *Languages* 2022, 7: 65. Disponible en: <https://www.mdpi.com/1533864>
- Martinez. y otros. (2025). Proyecto sectorial PS223LH 002-027 “Desarrollo de un sistema de certificación de dominio del Inglés para la educación superior cubana. Fase II” Título del programa: La formación integral universitaria para la innovación y el desarrollo sostenible. UCPEJV-MES.
- North, Brian (2023). The CEFR Companion Volume— What’s new and what might it imply for teaching/learning and for assessment? *CEFR JOURNAL—RESEARCH AND PRACTICE VOLUME 4* <https://doi.org/10.37546/JALTSIG.CEFR4-1>
- Rahman, M. The Advantages and Disadvantages of Using Qualitative and Quantitative Approaches and Methods in Language “Testing and Assessment” Research: A Literature Review. *Journal of Education and Learning*; Vol. 6, No. 1; 2017 pp 102-112 ISSN 1927-5250 E-ISSN 1927-5269 Published by Canadian Center of Science and Education. Disponible en: <https://files.eric.ed.gov/fulltext/EJ1120221.pdf>
- Rupp, André A., Vock, Miriam, Harsch, Claudia & Köller, Olaf. (2010) Technical Report Series: The Development, Calibration and Validation of Standards-based Tests. Institute for educational Progress. Humboldt University, Berlin, Germany.
- Hawkey, Roger. (2006) *Impact Theory and Practice Studies of the IELTS test and Progetto Lingue 2000*, Cambridge University Press. Disponible en: <https://www.cambridgeenglish.org/images/329230-studies-in-language-testing-volume-24.pdf>

### **Conflicto de intereses**

Los autores declaran que no existe conflicto de interés

### **Contribución autoral:**

**Marisol de la Caridad Patterson Peña:** *conceptualización, análisis formal, investigación, visualización, redacción del borrador original; redacción, revisión y edición.* Participó en el diseño y redacción del artículo, teniendo a su cargo el 55 % de la revisión bibliográfica,

procesamiento, recopilación y digitación de datos, así como análisis crítico y depuración de los mismos.

**Yoan Martínez Márquez:** *conceptualización, investigación, supervisión, revisión.* Brindó orientación y aportó perspectivas valiosas para la fundamentación teórica y el análisis preliminar de los datos. Contribuyó en el asesoramiento y la revisión del artículo en un 20 % del trabajo.

**Guillermo Manuel Negrín Ortiz:** *conceptualización, análisis formal, revisión.* Contribuyó con la adaptación de la encuesta aplicada y además con la revisión y aclaraciones relacionadas con los métodos y análisis estadísticos utilizados en un 15 % del trabajo.

**Claudia Harsch:** *conceptualización, supervisión, revisión.* Contribuyó en la revisión del artículo, aportando ideas y retroalimentación constructiva con respecto a los métodos estadísticos. Contribuyó además en la orientación de bibliografía relacionada con el uso del instrumento utilizado (la encuesta) en un 10 % del trabajo.