

ARTÍCULO

Evaluación de pensamiento crítico: Recuento de una experiencia interuniversitaria

Critical Thinking Assessment: An Interuniversity Experience Account

JULIÁN MARIÑO, NATALIA RONDEROS, JUAN JACCOBO GARZÓN MARTELO,
CARLOS CORTISSOZ, CAROLINA SEPÚLVEDA, JUANA HABER MARIÑO,
PAULINA YEPES VILLEGAS, JUAN SEBASTIÁN RUIZ, FELIPE BELTRÁN VEGA,
CLAUDIA YINETH CASTRO, JUANA MARÍA RÍOS JIMÉNEZ

jp.marino@uniandes.edu.co

Recibido el 13 de octubre del 2022; Aprobado el 10 de septiembre del 2023

RESUMEN

El desarrollo y promoción del pensamiento crítico resulta de gran importancia para las instituciones de educación superior, dado el valor que tienen para la vida académica y profesional de los estudiantes. Para explorar la posibilidad de contar con pruebas de desempeño que evalúen el pensamiento crítico, un grupo de siete universidades colombianas adaptó y aplicó una de estas pruebas a una muestra de 356 estudiantes. Este artículo tiene como objetivo hacer un recuento del proceso de adaptación, pilotaje y aplicación final de la prueba.

PALABRAS CLAVE: Pensamiento crítico; Evaluación; Pruebas de desempeño; Adaptación de pruebas; Educación superior

0185-2760/© 2016 Asociación Nacional de Universidades e Instituciones de Educación Superior A.C. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

ABSTRACT The development and promotion of critical thinking is of great importance for Higher Education Institutions, due to their implicit value for the academic and professional life of students. To explore the possibility of having performance tests that evaluate critical thinking, a group of seven Colombian universities adapted and applied one of these tests to a sample of 356 students. This article aims to make an account of the adaptation process, piloting and final application of the test.

KEYWORDS: Critical thinking; Assessment; Performance assessment; Test adaptation; Higher education

INTRODUCCIÓN

En el siglo XXI, tanto para la vida académica y profesional, como para el ejercicio ciudadano es fundamental el desarrollo del pensamiento crítico (PC). Éste es considerado como un conjunto complejo de habilidades que se manifiestan al momento de evaluar información para emitir juicios, en el marco de un propósito específico, utilizando la interpretación, el análisis, la evaluación y la inferencia, entre otros procesos de pensamiento de orden superior (Shavelson et al. 2018, p. 349). Desde esta perspectiva, es fundamental que las instituciones de educación superior integren en sus procesos formativos estrategias en las cuales los estudiantes desarrollen habilidades de conceptualización, síntesis, análisis y valoración de la información con el fin de promover el PC. El desarrollo del PC es uno de los objetivos más frecuentemente promovido por las instituciones de educación superior en Estados Unidos (Braun et al., 2020) y es obligatorio en las instituciones de educación superior en Colombia (MEN, 2006).

En este sentido, y para el contexto latinoamericano, Rolon (2014) plantea que uno de los objetivos de las universidades es proponer planes de formación y evaluación que permitan a los estudiantes tener espacios para poner en práctica sus habilidades de PC. En cuanto a los procesos evaluativos, en Colombia no se cuenta aún con una prueba que permita reflejar el desempeño de los estudiantes en esta habilidad. Es por eso que el propósito de este proyecto fue diseñar, adaptar, validar y aplicar una prueba a gran escala, en diferentes universidades colombianas, que evaluara el PC en estudiantes de distintos programas y semestres académicos. El objetivo fue desarrollar un instrumento que ofreciera información diagnóstica del PC de una muestra de estudiantes en las instituciones de educación superior participantes y estudiar la factibilidad de afinar una prueba interuniversitaria de desempeño en PC en el contexto colombiano.

El presente artículo relaciona, en sus distintas fases, el proceso de implementación de una prueba de PC, en el marco del proyecto IPAL (International Performance Assessment of Learning), con el grupo PALCO (Performance Assessment of Learning in Colombia) conformado por académicos de diferentes instituciones colombianas de educación superior: Universidad de los Andes, Universidad EAFIT, Universidad Jorge Tadeo Lozano, Universidad del Rosario, Corporación Universitaria Minuto de Dios, Universidad de Cundinamarca y Politécnico Grancolombiano.

Para articular la temática propuesta, este texto está compuesto por tres partes. En la primera, hacemos una conceptualización del PC, enfatizando sus principales características y la manera como lo abordamos para evaluarlo con la prueba de desempeño. En la segunda parte, hacemos referencia a los procesos de diseño, adaptación, pilotaje y aplicación de la prueba. En la tercera parte, exponemos cómo se llevó a cabo el proceso de codificación de las respuestas de 356 estudiantes.¹ Por último, exponemos los resultados generales del proyecto y finalizamos con algunas consideraciones y conclusiones.

MARCO TEÓRICO

En el diseño de la prueba partimos de la definición operativa de Shavelson et al. (2018, p. 341), según la cual el PC es una competencia multifacética que abarca “los procesos de conceptualización, análisis o síntesis, evaluación y aplicación de información con el objetivo de resolver un problema, decidir sobre un curso de acción, encontrar una respuesta a una pregunta compleja o sacar una conclusión”. Shavelson et al. (2018) recogen en su definición dos vertientes. Por un lado, la propuesta de Wheeler & Haertel (1993) que plantea que el PC no solo se expresa en contextos de formación académica, sino en los contextos más habituales de resolución de problemas. Por otro lado, la visión de Kosslyn & Nelson (2017) del PC en cuatro partes: (1) evaluación de afirmaciones, (2) análisis de inferencias, (3) evaluación de decisiones y (4) análisis de problemas. A partir de allí, Shavelson et al. (2018) construyen una definición que abarca los procesos observables en la ejecución de tareas que suscitan PC (conceptualización, análisis, evaluación, etc.) y que se dan naturalmente en la enorme variedad de problemáticas de la vida diaria.

Esta definición implica algunos compromisos con respecto a importantes cuestiones teóricas que aún están abiertas a discusión y que, si bien no podemos dirimir aquí, sí debemos reconocer y discutir brevemente. Estas cuestiones son: (1) ¿cuáles son los componentes del PC?, ¿es el PC un conglomerado de habilidades o disposiciones, o de ambas?; (2) ¿en qué tareas se expresa el PC?; (3) ¿cuál es la relación entre el PC y los conocimientos

¹ Este artículo, entonces, se ceñirá a las reflexiones que suscitó el proceso de elaboración y validación de la prueba, no a la exposición y análisis de sus resultados, los cuales serán objeto de otro artículo.

específicos?, ¿es la competencia en PC transferible de un dominio a otro, de un contexto disciplinar a otro?; y (4) ¿cómo puede evaluarse?

En lo que atañe a la cuestión de los componentes, podemos decir que esta definición, si bien es compatible con las aproximaciones que le otorgan gran importancia a las disposiciones, se enfoca exclusivamente en las habilidades cognitivas. Éstas incluyen no sólo los procesos cognitivos más generales (conceptualizar, analizar, sintetizar y evaluar), sino su expresión más concreta en tareas como detectar relevancia y consistencia, identificar agendas, comprender información cuantitativa y cualitativa, estructurar una posición, argumentarla y comunicarla de manera clara, entre otras (Case, 2005; Lipman, 1988). En el contexto de una prueba como esta, la evaluación de rasgos disposicionales (apertura mental, flexibilidad, deseo de informarse bien, razonabilidad, etc.) resulta difícil de materializar, en virtud de la enorme variedad de elementos que podrían evidenciarlos, así como del poco desarrollo de herramientas y recursos, probados empíricamente, para evaluarlos. Si bien podemos distinguir conceptualmente la *habilidad* de pensar críticamente de la *disposición* a hacerlo, discriminar los efectos de una y otra a través de una evaluación es particularmente difícil en la práctica (Norris, 1989).

En lo referente a las tareas, asumimos que el PC se evidencia en aquellas tareas de la vida diaria que implican la resolución de problemas, la toma de decisiones sobre un curso de acción, la respuesta a preguntas complejas y la inferencia de conclusiones, principalmente en contextos prácticos, de naturaleza social o política (Bonk & Smith, 1998; Fischer et al., 2009; Halpern, 1998; Silva, 2008; Shavelson, 2013). Este tipo de contextos no solo proveen la oportunidad de reflejar desempeños auténticos, sino que suelen ser contextos con vacíos de información y que tienen más de una interpretación o solución plausible o defendible. Además, suelen ser contextos que los estudiantes perciben como cercanos y auténticos.

En cuanto a la relación con conocimientos particulares, esta definición reconoce que el PC implica habilidades no específicas a un dominio en particular, aunque ciertamente requiere de alguno para aplicarse. El diseño e implementación de una prueba de PC para estudiantes universitarios de diversas carreras se compromete con la idea de que esta competencia, o al menos algunas de sus habilidades constitutivas, son genéricas y pueden ejercitarse independientemente del contexto de aplicación y de la formación disciplinar del estudiante. En este sentido, nuestro ejercicio, aunque no se compromete con la concepción del PC como exclusivamente genérico, sí se desmarca de una concepción para la cual pensar críticamente consiste en la apropiación de los estándares y valores de una cierta disciplina (cf. Paul, 1992). Si este fuera el caso, las habilidades de PC serían inherentes a sus dominios de acción o a las disciplinas subyacentes y, por ende, no podrían transferirse: pensar críticamente sería una cosa en las artes, otra en las matemáticas y otra enteramente distinta en las ciencias sociales. Nuestra prueba necesariamente observa habilidades como conceptualizar, sintetizar, analizar, entre otras, como independientes del universo de la prueba o de los dominios de experticia relevantes.

En cuanto a la evaluación de una competencia así concebida, es necesario enfocarla en el desempeño, en un escenario de simulación de alta fidelidad de situaciones cotidianas y reales como las que podrían encontrarse en medios de comunicación, redes sociales o, en general, en situaciones que hagan parte del debate público (Braun et al., 2020; Shavelson, 2013). Frente a esta simulación, el estudiante debe leer y comprender adecuadamente documentos que representan múltiples dimensiones, perspectivas y posiciones con respecto a la problemática. Luego, considerando la relevancia y la credibilidad, debe valorar la información cualitativa y cuantitativa sobre la cual construirá un juicio, propondrá una solución o recomendará un curso de acción.

ADAPTACIÓN Y DISEÑO DE LA PRUEBA

Frente al objetivo de evaluar el PC en las universidades colombianas, se optó por adaptar un escenario de desempeño ya existente. La razón de esta elección, y de no desarrollar una nueva prueba, se relacionó con los costos económicos y el tiempo que debían ser invertidos en el proceso (Schendel & Tolmie, 2017). Dado que el objetivo del proyecto era contar con un instrumento que evaluara el PC en estudiantes colombianos de educación superior, y no reproducir una prueba existente ni establecer una comparación entre el desempeño de éstos y los estudiantes de otro país, desde el inicio fue claro que el instrumento evaluativo debería ser sensible al contexto colombiano y responder a sus especificidades sociales, económicas, políticas y culturales² (Ronderos et al., 2021).

El primer paso fue la exploración de escenarios elaborados por diferentes grupos del proyecto IPAL. Se revisaron tres opciones: (i) la admisión universitaria por legado, en donde los estudiantes deben analizar los perfiles académicos de tres candidatos y los vínculos con la universidad, para así recomendar quién debe ser admitido (Estados Unidos) (Braun et al., 2019); (ii) la crisis de refugiados, en la que los evaluados deben elaborar una recomendación sobre aumentar o no el número de refugiados que acoge un país (Hyytinen & Toom, 2019); (iii) la construcción de un parque eólico, en la cual los estudiantes deben hacer una recomendación que ayude a decidir si aceptan o no la oferta de una compañía para su construcción en una comunidad (Alemania) (Shavelson et al., 2019). Luego de analizar la relevancia y pertinencia de las temáticas de las pruebas para los estudiantes colombianos, decidimos adaptar el examen de la Universidad de Helsinki³, *Crisis de refugiados*.

² Las pruebas de desempeño, a diferencia de las pruebas estandarizadas basadas en preguntas de selección múltiple, enfrentan un mayor reto de ser sensibles al contexto local. Más sobre lineamientos para adaptar pruebas de desempeño de pensamiento crítico puede encontrarse en Ronderos et al (2021).

³ La prueba es de autoría de los profesores Heidi Hyytinen y Auli Toom y fue diseñada para estudiantes de pregrado.

Este escenario trata de un país ficticio europeo que enfrenta un aumento en el número de refugiados que llegan a su territorio. El gobierno debe decidir entonces si aumentar o no el número de refugiados acogidos, teniendo en cuenta las posiciones que exponen diferentes actores. Así, los estudiantes deben enunciar las perspectivas que están a favor y en contra de que estos centros aumenten y estructurar una propuesta de acción sobre esta problemática, basados en la información de los documentos presentados en la prueba (Hyytinen & Toom, 2019; Ronderos et al., 2021).

Luego de la lectura en grupos de trabajo, consideramos que el tema de este escenario suponía un reto interesante para los estudiantes de las instituciones involucradas en el proyecto puesto que, cuando debatíamos esas alternativas, el tema migratorio estaba muy presente en los medios y el debate público nacional. Concretamente, dadas las condiciones políticas, económicas y sociales del país vecino, Venezuela, un número significativo de personas seguía migrando a diferentes países de América Latina. Al compartir su frontera terrestre, Colombia era uno de los principales destinos de la población venezolana migrante. A diario, desde diferentes ámbitos, los medios de comunicación se ocupaban de esta problemática, lo cual ha generado opiniones diversas, controversias, circulación de información falsa, y análisis desde diferentes orillas.

La adaptación local con base en el escenario finlandés se hizo en cinco fases antes de su aplicación final, las cuales se detallan a continuación. Este proceso comprendió la (i) definición del constructo, (ii) traducción, (iii) contextualización del tema de refugiados, (iv) selección de documentos y desarrollo de las tareas o retos (v) aplicación del ejercicio de pensamiento en voz alta. Posteriormente, elaboramos la prueba definitiva, reuniendo las observaciones derivadas de la etapa anterior. Además, realizamos un pilotaje de la prueba que permitiera reunir otras percepciones y trabajar en la construcción de la rúbrica y la aplicación final.

Adaptación de la prueba

Definición del constructo. El primer paso del proyecto fue definir qué facetas del PC se querían evaluar con la prueba. Después de varias deliberaciones, el grupo optó por las siguientes facetas: (1) análisis y evaluación de información y argumentos; (2) comunicación coherente de una posición justificada en evidencia; (3) reconocimiento y consideración de las posibles consecuencias de un curso de acción; y (4) razonamiento cuantitativo. Esta última se entiende como el análisis y evaluación de información cuantitativa para apoyar la toma de decisiones, es decir, se aborda desde el PC. Esta faceta se incluyó por la cantidad de información cuantitativa que tenía la prueba finlandesa original, y por ser el razonamiento cuantitativo una de las competencias genéricas más reconocidas en la educación superior colombiana.

Traducción. El escenario original estaba en finlandés. Antes de iniciar la adaptación del instrumento, y con el objetivo de poder entenderlo, se realizó una traducción al español. La traducción presentó una serie de retos importantes. En primer lugar, al no tener en el equipo nadie que hablase finlandés, fue necesario explorar el sentido del texto original por medio de diversas traducciones automáticas al español y al inglés. En segundo lugar, verificamos que la traducción no se limitara a equivalencias lingüísticas, sino que se enfocara en la construcción del sentido global y la coherencia interna de la prueba.

Así pues, primero hicimos una traducción automática del finlandés al español usando la herramienta de traducción de Google; no obstante, generó inquietudes con respecto a la consistencia de algunos pasajes. En esa medida, optamos por hacer otras traducciones (del finlandés al inglés y del inglés al español) empleando herramientas de Google y Microsoft. Posteriormente, contrastamos estas traducciones y pasamos a la edición. En esta etapa, identificamos algunas palabras y expresiones cuyos significados y usos podrían ser ajenos a la situación o al contexto de migración en la región, razón por la cual las cambiamos por otras más cercanas, que permitieran apropiarse de manera más directa de la problemática. Tal es el caso del término *refugiados*, el cual analizaremos en el siguiente apartado. Finalmente, después de obtener una prueba consistente en español, contactamos a los autores de la prueba original y a un grupo de expertos de IPAL, con el fin de recibir retroalimentación y constatar que la prueba se ajustara al marco de IPAL.

Contextualización del tema de refugiados. La revisión de la prueba nos llevó a determinar un tópico más relacionado con la realidad de los estudiantes de las universidades participantes, de modo que decidimos centrar la problemática en la migración, en contraste con el tema original que era el de refugiados. La elección del tema de la prueba suscitó discusiones sobre la caracterización del concepto de refugiado en contraposición al de migrante. Dichas discusiones fueron clave en su traducción contextual, pues la situación del país en el que se diseñó la prueba, Finlandia, y su problemática, tiene unas características geopolíticas particulares que no se aplican al contexto latinoamericano.

Según la oficina del Alto Comisionado de Naciones Unidas para los Refugiados, un refugiado es aquella persona que se encuentra fuera de su país de origen y que, al no poder regresar —por cuestiones de persecución, conflicto o violencia— requiere protección internacional (ACNUR, 2022). Por otro lado, un migrante es aquella persona que se está trasladando o se ha trasladado a través de una frontera internacional o dentro de un mismo Estado, fuera de su habitual lugar de residencia, independientemente de su estatus legal, de si el traslado es voluntario o involuntario, y de su causa (Naciones Unidas, 2022a). El grupo de trabajo consideró que en Colombia no se utiliza tan frecuentemente en el debate público el concepto de refugiado, es más recurrente el uso del concepto de migrante, como se puede constatar a través de sus medios de comunicación, discusiones y debates sobre el tema. Adicionalmente el concepto de migrante refleja de mejor manera la realidad migratoria latinoamericana.

Selección de documentos. El primer paso para la selección de documentos fue identificar aquellos documentos de la prueba original que eran disonantes con nuestro contexto, y desarrollar una búsqueda de textos y noticias que tuvieran relación con la migración en nuestra región. El objetivo era obtener información y casos que fueran cercanos a la realidad latinoamericana para ubicar la prueba dentro de un contexto similar al nuestro. Los textos elegidos analizaban la migración desde diferentes ámbitos (económico, jurídico, humanitario, social, cultural) y presentaban posiciones a favor y en contra de la migración. Asimismo, incluían datos e información, en algunos casos sesgada, y exigían un análisis tanto de las fuentes (autores y medios de publicación) como de las evidencias que citaban, para determinar su credibilidad y veracidad.

Encontramos 26 documentos. Para la selección de los que se utilizaron en las etapas posteriores de validación, elaboramos una tabla en la que se presentaba, para cada uno, su categorización en términos del constructo de la prueba (confiabilidad, relevancia, implicaciones éticas y presencia de información cuantitativa) y sus características (autenticidad en el contexto latinoamericano, complejidad del texto, posición frente a la migración, ideas principales). A partir de estos aspectos, analizamos las particularidades de cada uno de los documentos para determinar cuáles de ellos se ajustaban mejor a la naturaleza de la prueba y realizar así el proceso de adaptación. En la Tabla 1 exponemos los 8 textos escogidos inicialmente para las primeras versiones de la prueba de PC (A, B y C).

Tabla 1. Textos para versiones A, B y C

Texto	Descripción
1. Correo electrónico de Patricio Ríos, FT, Ph.D. Director Internacional de la Casa del Migrante	Documento adaptado de los documentos de la prueba original. Formato de correo electrónico proveniente del director de la Casa del Migrante, quien ofrece sus servicios y expone los beneficios de su programa para aliviar la situación de los migrantes.
2. Extracto del Informe sobre migración de la Fundación para el Futuro (FF)	Documento nuevo, para reemplazar un documento similar, con datos cuantitativos sobre la relación entre migración y criminalidad, que estaba situado en el contexto europeo. Documento adaptado de informe real de datos de migración y criminalidad en Colombia, de la Fundación Ideas para la Paz (Castillo et al., 2019).
3. Extracto de reporte de la OIM: <i>Las migraciones en América Latina y el Caribe. Una mirada de género</i>	Documento nuevo. Artículo de la OIM que expone la problemática migratoria en América Latina desde una perspectiva de género (Fries, 2019).
4. La influencia negativa de la migración en el proceso educativo de la ciudad	Documento nuevo, para reemplazar un artículo académico de la prueba original que resultaba muy complejo y poco adecuado para el contexto colombiano. Adaptado de un extracto de texto académico sobre las consecuencias de la migración en el proceso educativo (Ramón & Guahichullca, 2018).
5. Artículo de periódico de <i>El Dorado Justo</i>	Documento adaptado de la prueba original. Artículo de opinión sobre la criminalidad y la migración.

Tabla 1. Continúa

Texto	Descripción
6. Seguridad ciudadana y migración	Documento nuevo, para reemplazar un artículo académico de la prueba original que resultaba muy complejo y poco adecuado para el contexto colombiano. Adaptado a partir de un informe de la Fundación Ideas para la Paz, que expone causantes y consecuencias tanto positivas como negativas de la migración (Castillo et al., 2019).
7. Extracto de la Declaración Universal de Derechos Humanos	Documento de la prueba original, tomando la versión en Español de las Naciones Unidas. Extracto con sus respectivos artículos.
8. Extracto del pacto Internacional de derechos civiles y políticos	Documento nuevo, para complementar los derechos humanos. Pacto de derechos civiles con sus respectivos artículos (Naciones Unidas, 2022b).
9. Comunicado de Gonzalo Harboe, senador de la República del Dorado	Documento nuevo. Artículo adaptado con información económica a favor de la migración proveniente de un senador de la república (Harboe, 2018).
10. Extracto adaptado de la página web oficial de la Presidencia de la República del Dorado	Documento nuevo, adaptado de la página de la presidencia de EEUU durante el gobierno de Donald Trump sobre la migración en este país (White House, 2019).

Fuente: Elaboración propia.

Tarea o reto. Siguiendo el marco IPAL, para observar el desempeño en PC, utilizamos una combinación de preguntas abiertas, en forma de texto argumentativo, y cerradas, de respuesta de selección múltiple. Además, en todas las versiones se expuso de la misma manera la problemática que el país estaba viviendo, así como las instrucciones y cinco preguntas con base en los textos. Estas fueron las preguntas planteadas:

1. Con respecto a la relación entre migración y criminalidad, ¿qué conclusiones se pueden extraer de las cifras presentadas?
2. ¿Qué información y perspectivas de los documentos apoyan establecer una política de apertura hacia los migrantes? Justifique su respuesta teniendo en cuenta el contenido de los documentos y su confiabilidad.
3. ¿Qué información y perspectivas de los documentos apoyan restringir la entrada de migrantes al país? Justifique su respuesta teniendo en cuenta el contenido de los documentos y su confiabilidad.
4. (Versión A) Como miembro del grupo de inmigración, elabore un texto que responda a la pregunta: ¿qué debería hacer el gobierno de la República de El Dorado ante la expectativa de que la presión migratoria siga creciendo en los próximos años? Por favor, dé las razones de sus recomendaciones y haga referencia a la información de los documentos que apoya su conclusión. Justifique su respuesta cuidadosamente.

(Versión B y C) Como miembro del grupo de inmigración, elabore un texto que responda a la pregunta: ¿El gobierno de la República del Dorado debería restringir la entrada de migrantes al país? Por favor dé las razones de sus recomendaciones y haga

referencia a la información de los documentos que apoya su conclusión. Justifique su respuesta cuidadosamente.

5. ¿Qué información adicional podría necesitar el Comité? ¿Por qué sería útil esta información?

La diferencia entre las versiones de las pruebas radicaba en dos elementos: los textos que las conformaban y la formulación de la cuarta pregunta. En todas las versiones se incluyeron los textos presentados en la tabla, con las siguientes variaciones. La versión A presentaba el texto de la OIM y solamente un texto que hacía referencia a los derechos humanos (texto 7, ver Tabla 1). Por otro lado, en la versión B cambiamos el texto de la OIM por un texto académico sobre las consecuencias de la migración en el proceso educativo (texto 4, ver Tabla 1) y dos textos relacionados con los derechos humanos y civiles (textos 7 y 8, ver Tabla 1). Por último, la versión C presentaba el texto de la OIM y los dos textos sobre derechos humanos y civiles. En cuanto a las preguntas, en la versión A la cuarta pregunta indagaba por lo que se debería hacer ante la problemática migratoria; y en las versiones B y C se preguntaba si se debía restringir o no la migración. Fue así como, con estas tres pruebas, realizamos el proceso de validación descrito en la siguiente sección.

Pensamiento en voz alta y entrevistas (laboratorios cognitivos). Para la validación de la prueba, se utilizó la técnica de pensamiento en voz alta, ampliamente utilizada por la psicología cognitiva para analizar actividades de resolución de problemas y procesos de escritura. Esta metodología “implica el uso de sujetos ‘pensando en voz alta’ de manera continua mientras llevan a cabo una actividad” (Castells, 2007, p. 29).

Con los responsables de realizar los ejercicios de pensamiento en voz alta se acordaron protocolos que especificaban en detalle cómo se debían adelantar. De acuerdo con esas indicaciones, se debía solicitar a los estudiantes que, al tiempo que respondían cada una de las preguntas de la prueba, verbalizaran sus procesos cognitivos, desde el momento en que iniciaban la lectura, pasando por la reflexión sobre las respuestas, hasta el momento de su escritura. Para aplicar el protocolo de pensamiento en voz alta, cada universidad eligió entre 3 y 5 estudiantes, con los cuales se realizó el ejercicio durante tres horas aproximadamente.

El protocolo se aplicó individualmente y fue asistido por un facilitador, encargado de orientar a los participantes, y un revisor, a cargo de registrar las observaciones. En total participaron 23 estudiantes. Después del desarrollo de la prueba, los responsables de la aplicación, en una entrevista estructurada, solicitaron a cada estudiante su opinión sobre: la precisión de las instrucciones; la credibilidad, relevancia, claridad, interés y realismo de las fuentes de información; el nivel de dificultad de la prueba; el tiempo asignado para resolverla; el interés que despertaba el caso planteado, es decir, si lo conectaba con la prueba y lo motivaba a responderla; la percepción de verosimilitud de la situación planteada; y la influencia de creencias, experiencias y conocimientos previos.

La prueba fue clara y precisa para el 87% de los estudiantes. El restante 13% consideró que, aunque la problemática era clara, se necesitaban instrucciones más detalladas para elaborar una respuesta. Con respecto al nivel de dificultad de la prueba, los estudiantes la calificaron, en promedio, con 2.8, en una escala de 1 a 5 (donde 1 es muy fácil y 5 muy difícil). De los 23 estudiantes, 3 la consideraron muy fácil, pero ninguno muy difícil. Además, 14 de ellos manifestaron que la prueba podría ser más fácil para estudiantes de áreas de ciencias sociales y humanas. En cuanto al tiempo de respuesta, solo 7 estudiantes plantearon que necesitarían más tiempo. En los aspectos motivacionales, el interés hacia la prueba se situó en 4.6 y hacia el tema en 4.8, en una escala de 1 a 5 (donde 1 es nada interesante y 5 muy interesante). En cuanto a la verosimilitud, la percepción de realismo de la prueba fue 4.9 y la de cercanía al contexto 4.5. Sin embargo, un grupo de estudiantes consideró que algunas fuentes de información no eran del todo realistas, ya fuese por el formato (tipografía, color, diseño), el contenido (nombres, lugares, situaciones), el sesgo muy evidente en algunas voces, la relevancia de algunos documentos o la credibilidad de los datos.

Adicionalmente, el 56,5 % de los estudiantes respondió que sus creencias personales sí influyeron en la toma de decisión frente a la situación, aunque seis estudiantes afirmaron haber decidido en contra de sus creencias. Además, el 90 % dijo haber considerado aspectos éticos para elaborar su respuesta.

Por último, la aplicación del protocolo de pensamiento en voz alta permitió revisar el funcionamiento de la prueba, así como analizar la articulación de la rúbrica: coherencia en su estructura, precisión de los criterios y diferenciación o gradación de los niveles de desempeño.

Conformación de las versiones de la prueba para la aplicación piloto. Al observar que el desempeño en esta prueba dependía crucialmente de interpretar y analizar correctamente la información cuantitativa y, por otro lado, de reconocer deberes y obligaciones de naturaleza ética, decidimos hacer tres nuevas versiones basados en esta consideración: una versión con un énfasis cuantitativo, otra con un énfasis ético y una versión mixta que mezclaba elementos de las anteriores dos. Este ejercicio nos llevó a descartar el uso del texto 3 (de la OIM) y a utilizar únicamente el 4 (referente al proceso educativo) (ver Tabla 1). Por otra parte, el nombre del país cambió de El Dorado a Valverde, dado que el nombre inicial evocaba a Colombia. Finalmente, añadimos algunas preguntas cerradas al finalizar la prueba, con el fin de obtener retroalimentación que permitiera conocer las impresiones personales de los estudiantes sobre la misma.

Para la versión cuantitativa, eliminamos uno de los textos de derechos humanos, texto 8, manteniendo el 7, y planteamos solo dos preguntas, una que le pedía al estudiante analizar la relación entre migración y criminalidad, basándose en los textos que presentaban información cuantitativa, y otra que le pedía establecer su posición sobre si restringir o no la migración. En esta prueba, el estudiante tomaría su decisión basado principalmente en

el análisis de datos. En la versión ética, acortamos el texto 2 con información cuantitativa sobre criminalidad y migración y dejamos los dos textos sobre derechos humanos, 7 y 8 (ver Tabla 1). Además, formulamos una sola pregunta en la que se le pedía al estudiante que expusiera su posición sobre si restringir o no la migración, enfatizando en el impacto que su decisión podría generar en millones de personas. Por último, en la versión mixta, incluimos todos los textos y empleamos la primera pregunta de la versión cuantitativa y la pregunta de la versión ética.

Asimismo, incluimos preguntas de selección múltiple en las tres versiones, que evaluaban la comprensión de la información. Con estas preguntas buscamos controlar el nivel de lectura de los estudiantes. Estas preguntas se presentaban una vez el estudiante respondía todas las abiertas.

Aplicación piloto

Con las tres versiones de la prueba (énfasis ético, cuantitativo y mixto) se realizó la aplicación piloto, mediante un cuestionario en línea cargado en la plataforma de exámenes del Centro de Evaluación de la Universidad de los Andes, con un tiempo límite para su respuesta de 1 hora y 45 minutos. En esta aplicación participaron 154 estudiantes de diferentes regiones del país y de distintas áreas del conocimiento como administración, educación, ciencias sociales e ingeniería, provenientes de las 7 universidades colombianas.

Al finalizar la aplicación piloto, el proceso de codificación de las respuestas contó con 33 codificadores que recibieron un entrenamiento inicial en el que se utilizaron 10 de las respuestas obtenidas. Una vez los criterios de evaluación fueron claros, se realizaron codificaciones dobles (dos personas evaluaban una misma respuesta) y 10 personas se encargaron de revisar las calificaciones de aquellas repuestas para las que las primeras codificaciones presentaran diferencias grandes entre sí. El proceso de codificación se realizó por medio de la Plataforma del Centro de Evaluación Educativa de la Universidad de los Andes.

Una vez realizado lo anterior, obtuvimos los resultados de los pilotajes. A partir de las respuestas dadas por los estudiantes, llegamos a resultados similares entre las versiones aplicadas, por lo que dedujimos que los distintos énfasis de la prueba no lograban evidenciar diferencias notables en los análisis y respuestas de los participantes. Asimismo, según la información recolectada, obtuvimos poca evidencia de un análisis de los documentos y un razonamiento ético y cuantitativo profundos por parte de los estudiantes, y poca capacidad de discriminación entre la confiabilidad y validez de los documentos. En términos de la generalizabilidad (Shavelson & Web, 1991) los resultados de los 16 criterios

codificados en la rúbrica oscilaron entre 0.60 y 0.82, con un promedio de 0.72. Asimismo, 6 de estos 16 criterios obtuvieron puntajes menores a 0.60 con notables dificultades en sus definiciones.

Por último, se aplicó una encuesta de retroalimentación sobre las diferentes versiones de la prueba. En sus respuestas, encontramos que, en una escala de 1 a 5, los estudiantes consideraron que las pruebas tenían un nivel de dificultad medio, con puntajes oscilando entre 3.4 y 3.5. Por otro lado, la mayoría de los estudiantes (entre 68% y 71%) reportaron que sus creencias previas sobre la migración tuvieron un rol importante al momento de dar respuesta a las preguntas abiertas, y una gran mayoría (entre 88% y 92%) reportó que sus respuestas estaban alineadas con las creencias que ya tenían previamente, lo cual era un resultado esperable antes de iniciar la fase de pilotaje, pues ya lo habíamos notado también en el ejercicio de pensamiento en voz alta.

Versiones definitivas de la prueba

La aplicación piloto nos permitió observar que la prueba se tornaba demasiado larga y que el tiempo de lectura y análisis de los documentos casi siempre sobrepasaba el tiempo de elaboración de la respuesta. Como resultado de lo anterior, optamos por crear, a partir de la versión mixta, dos versiones, una larga y otra corta, con el fin de reducir la carga de lectura para los estudiantes (ver Tabla 2). La versión larga incluyó todos los documentos mientras que la versión corta tuvo una compilación de la información de varios documentos en uno solo. Este documento fue presentado en el formato visual de la plataforma Twitter y allí se consignaron las ideas principales de los textos: 1, 4, 9 y 10 (ver Tabla 1). Para las dos versiones se formularon 4 preguntas:

1. Con respecto a la relación entre migración y criminalidad, ¿qué conclusiones se pueden extraer de las cifras presentadas?
2. ¿Qué información y perspectivas de los documentos apoyan establecer una política humanitaria de apertura hacia los migrantes?
3. ¿Qué información y perspectivas de los documentos apoyan restringir la entrada de migrantes al país?
4. ¿Usted recomienda restringir la entrada de migrantes al país? ¿Por qué?

Estas dos versiones fueron las utilizadas en la aplicación final en las universidades. Además, con el propósito de obtener más información sobre el nivel de comprensión de lectura, análisis crítico y razonamiento cuantitativo de los estudiantes, se incluyeron en las versiones finales las preguntas de selección múltiple y respuesta corta.

Tabla 2. Documentos de las versiones larga y corta de la prueba

Versión larga	<ol style="list-style-type: none"> 1. Correo electrónico de Patricio Ríos, PhD. Director Internacional de la Casa del Migrante 2. Extracto del Informe sobre migración de la Fundación para el Futuro 3. Extracto de investigación: La influencia negativa de la migración en el proceso educativo de la ciudad 4. Artículo de periódico de Valverde ONLINE 5. Seguridad ciudadana y migración 6. Extracto de la Declaración Universal de Derechos Humanos 7. Comunicado de Gonzalo Harboe, profesor emérito de la UNAV de Valverde 8. Página web oficial de la Presidencia de la República de Valverde
Versión corta	<ol style="list-style-type: none"> 1. Extracto del Informe sobre migración de la Fundación para el Futuro 2. Debate en Twitter sobre la #InmigraciónenValverde 3. Artículo de periódico de Valverde ONLINE 4. Seguridad ciudadana y migración 5. Extracto de la Declaración Universal de Derechos Humanos

Fuente: Elaboración propia.

Elaboración de la rúbrica

Una vez obtenidas las primeras respuestas en el piloto de la prueba se inició el trabajo para poner a punto la rúbrica de evaluación. Cabe aclarar que fueron numerosas las modificaciones que sufrió este instrumento a lo largo del proyecto, por lo que aquí se expondrán únicamente las variaciones entre la primera versión y la última. Dado que el escenario inicial de Finlandia, que utilizamos como base para construir la nuestra, no contaba con una rúbrica de evaluación, adoptamos como punto de partida las rúbricas de evaluación de PC previamente utilizadas por el Centro de Evaluación de la Universidad de los Andes. Esta rúbrica contaba con 4 dimensiones, cada una de las cuales incluía una serie de criterios, para un total de 34 criterios. Las cuatro dimensiones eran: (1) explorar y evaluar; (2) análisis de perspectivas; (3) elaborar propuestas de solución; y (4) comunicación. Como se señaló más arriba, por la definición del constructo adoptada para el proyecto, descartamos la dimensión de análisis de perspectivas al inicio del proyecto. Por otra parte, para dar cuenta de las dimensiones correspondientes a los énfasis de las nuevas pruebas, debimos desarrollar dos dimensiones nuevas: de razonamiento ético y de razonamiento cuantitativo.

La primera dimensión, *Explorar y evaluar*, incluía criterios que evaluaban tanto el nivel de comprensión de la información de los textos en la prueba como la identificación de argumentos y su evaluación en términos de credibilidad y confiabilidad. En la versión final esta dimensión quedó constituida por dos criterios: 1) comprensión de los documentos, en donde se evalúa la comprensión de las ideas principales y secundarias; y 2) identificación y abordaje de las dimensiones presentes en los textos, como jurídica, económica, laboral, criminal y social. Con esto aseguramos que los estudiantes se basaran en los textos proporcionados en la prueba para poder construir su respuesta y contestar a las preguntas presentadas.

La dimensión *Elaborar propuestas de solución*, tenía como propósito evaluar la propuesta del estudiante. Aquí considerábamos la calidad de su argumentación y contraargumentación, así como el uso de las fuentes disponibles. Luego de las múltiples modificaciones llevadas a cabo en las diferentes fases del proyecto, quedó constituida por tres criterios: 1) la presentación de una tesis sobre la posición del estudiante, 2) la sustentación de la posición del estudiante con argumentos coherentes y bien articulados; y 3) el uso de evidencia relevante y el cuestionamiento de inconsistencias, supuestos o credibilidad de las fuentes. De esta manera quedaron incluidos en esta dimensión aspectos relacionados con argumentación y credibilidad de las fuentes que previamente hacían parte de *Explorar y evaluar*.

La dimensión de *Comunicación* evaluaba la producción textual en términos de claridad, concisión y respeto por las reglas ortográficas y gramaticales del español. A lo largo del proyecto, esta dimensión tuvo también algunos cambios, pero su esencia se mantuvo, dado que su propósito siempre fue evaluar los textos en términos de escritura y claridad de las distintas ideas. Finalmente, quedó constituida por tres criterios: 1) estructura y articulación lógica de las ideas, entre y dentro de los párrafos; 2) mención de títulos y autores cuando se hace referencia a los documentos; y 3) buen uso del vocabulario, la ortografía y la gramática.

Respecto a la una nueva dimensión de *Razonamiento cuantitativo*, en un inicio partimos de cuatro criterios: comprensión, uso, análisis y abordaje de los datos. No obstante, en esas primeras definiciones, las descripciones de los criterios se traslapaban entre sí, lo cual hacía difícil su discriminación y dificultaba su evaluación. Es por eso que, durante el proceso de consolidación de la dimensión, unimos algunas descripciones y simplificamos definiciones, lo que finalmente nos condujo a dejar solo dos criterios: 1) interpretación de la información cuantitativa, que hace referencia a la manera en que el estudiante interpreta la naturaleza de las variables presentadas, es decir, si en su respuesta identifica correctamente a qué variables se refieren las distintas cifras que se presentan en los documentos; y 2) análisis cuantitativo, enfocado en la correcta lectura de los datos cuantitativos y la formulación de consecuencias que de ellos pueden obtener. Lo anterior comprende la capacidad del estudiante para leer gráficas, identificar tendencias, relacionar variables y realizar inferencias a partir de su análisis.

Por otra parte, exploramos el potencial de la prueba para examinar competencias de razonamiento ético. Aunque el interés por observar esta dimensión estaba ya presente en la implementación de la prueba original, la exploración de varios énfasis (cuantitativo, ético y mixto) nos permitió ver que el razonamiento ético juega un rol muy importante. Sin embargo, los intentos iniciales por evaluar esta dimensión no arrojaron buenos resultados en términos de validez y confiabilidad, por lo que se prescindió de incluir esta dimensión en la rúbrica. Aun así, cabe aclarar que los esfuerzos por crear un mecanismo que evalúe esta dimensión siguen presentes, por lo que se espera exponer el proceso y los resultados en detalle en una próxima publicación.

Después de los análisis y reflexiones antes presentados, llegamos a una versión definitiva de la rúbrica, que incluía cuatro dimensiones con diez criterios en total, presentados en la tabla 3. Cada criterio tenía 5 niveles, con descriptores elaborados para el nivel más bajo (1), el intermedio (3) y el más alto (5).

Tabla 3. Dimensiones y criterios de evaluación de la rúbrica

Dimensiones	Criterios
Explorar y evaluar	<ul style="list-style-type: none"> • Comprensión de los documentos (excluida la información cuantitativa) y de las posiciones que presentan frente a la migración. • Riqueza del análisis (dimensiones o aspectos abordados).
Elaborar propuesta de solución	<ul style="list-style-type: none"> • Presenta una tesis del análisis de criminalidad y migración y una posición con respecto a si restringir o no la migración. • Sustenta su posición y su análisis con argumentos coherentes. • Uso de evidencia relevante y cuestionamiento de afirmaciones, supuestos, inconsistencias y credibilidad de las fuentes.
Razonamiento cuantitativo	<ul style="list-style-type: none"> • Interpretación de la información cuantitativa. • Análisis cuantitativo.
Escritura y comunicación eficaz	<ul style="list-style-type: none"> • Efectividad y claridad de la comunicación. • Referencia a los documentos. • Vocabulario, ortografía y gramática.

Fuente: Elaboración propia.

FASE DE APLICACIÓN

La aplicación de la prueba se llevó a cabo durante los meses de noviembre y diciembre de 2020. En total participaron 356 estudiantes a quienes les fueron asignadas aleatoriamente las dos versiones, buscando que dos tercios respondieran la versión larga y un tercio la corta (ver Tabla 4). Los participantes fueron estudiantes de las 7 universidades, inscritos en programas académicos de diferentes áreas de conocimiento: ciencias sociales y humanas, ciencias naturales, exactas y aplicadas, artes, entre otras, y de distintos semestres (entre 1º y 12º), incluyendo algunos aspirantes a maestrías.

Tabla 4. Estudiantes por universidad que respondieron cada versión de la prueba

Universidad	Versión larga	Versión corta	Mujeres	Hombres
Universidad EAFIT	24	11	20	15
Universidad Jorge Tadeo Lozano	35	23	40	18
Universidad del Rosario	39	20	38	21
Universidad de Cundinamarca	32	17	28	21
Corporación Universitaria Uniminuto	52	22	51	23
Politécnico Grancolombiano	13	6	12	7
Universidad de los Andes	31	25	19	37
Total	226	124	208	142

Nota: Del total, seis estudiantes no se pudieron clasificar dentro de las versiones larga y corta, debido a que no contestaron la sección de preguntas cerradas de la prueba.

Tabla 5. Estudiantes por semestre y área que respondieron la prueba

Facultad	Semestre			Total
	Bajos	Medios	Altos	
Administración	14	58	50	122
Arquitectura & Artes	10	30	2	42
Sociales & Afines	27	39	49	115
STEM	19	32	25	76
Total	70	159	126	355

Nota: La variable semestre agrupa los semestres 1, 2 y 3 como "bajos", los semestres 4, 5 y 6 como "medios" y los semestres 7, 8, 9 y 10 como "altos". El caso faltante, que no tiene facultad, está en semestres altos y pertenece a la universidad Jorge Tadeo Lozano.

Codificación

La codificación de la aplicación final se realizó durante los meses de febrero y marzo de 2021. Los codificadores fueron asistentes y profesores participantes de las 7 universidades y todos recibieron un entrenamiento durante agosto y septiembre de 2020. Para esto, elegimos respuestas diversas que fueron calificadas por todos los codificadores. Después de un ejercicio individual de evaluación, analizamos los acuerdos y las discrepancias. Esta dinámica nos permitió identificar qué criterios se tornaban confusos o no se ajustaban a algunas respuestas, por lo que resultó necesario redactarlos de manera más clara o precisa y llegar a consensos sobre cómo calificarlos.

Por otra parte, con el fin de apoyar la toma de decisiones de los codificadores, durante la etapa de entrenamiento elaboramos un banco de ejemplos de respuestas para cada criterio. Los ejemplos ilustraban tipos de respuesta que se ajustaban al nivel más alto de los criterios o exponían errores de razonamiento y análisis que las ubicaban en los niveles más bajos. Asimismo, retomamos los documentos de la prueba y determinamos aspectos clave para su comprensión, por ejemplo: tema, ideas principales y secundarias, tesis y propósito comunicativo. También enfatizamos en los ámbitos desde los cuales se abarcaba la problemática de la migración, las posiciones que se adoptaban frente a restringirla o no y los problemas que se presentaban (sesgos, falta de autoridad, inconsistencias, información en conflicto, falta de evidencias, etc.).

Después del entrenamiento, el proceso de codificación se desarrolló en dos momentos. En el primero, una misma prueba fue evaluada por dos codificadores. Además de marcar el nivel de desempeño alcanzado por un estudiante en cada criterio, los codificadores, en la mayoría de los casos, escribían sus justificaciones y llamaban la atención sobre aspectos particulares de la respuesta que determinaban su calificación. Luego de revisar estas codificaciones y de llegar a acuerdos sobre las discrepancias que se presentaron, dimos paso al segundo momento. En este, se vinculó un tercer codificador al proceso, quien ejerció el papel de revisor. Su función principal consistió en mediar y consensuar razonadamente

las calificaciones que presentaban discrepancias de más de dos puntos, apoyándose en los acuerdos a los que se llegaron en las reuniones previas. De esta manera, el proceso de codificación contó con tres lectores por respuesta según el nivel de discrepancia, lo cual permitió la imparcialidad en la calificación y una mayor unidad de criterio en la aplicación de la rúbrica.

CONCLUSIÓN

Los cambios y las exigencias del mundo actual imponen a las instituciones de educación superior el reto de lograr desarrollar el PC entre sus estudiantes, en la medida en que demandan profesionales que cuenten con la capacidad de evaluar información para tomar decisiones, argumentarlas y sustentarlas. Para orientar los esfuerzos que contribuyan a consolidar estas habilidades complejas, y no solo sus componentes de forma independiente, se requiere evaluar el desempeño de los estudiantes en situaciones contextualizadas, en escenarios realistas, con desafíos que involucren decidir sobre un curso de acción, responder a una pregunta compleja o resolver un problema definido solo parcialmente.

Desde esta perspectiva, es evidente el interés y la utilidad de la experiencia de diseñar, aplicar y validar una prueba que reflejara, de manera confiable, el desempeño de esta competencia en estudiantes colombianos de diferentes instituciones de educación superior. En términos del objetivo planteado, de explorar la posibilidad de tener pruebas de desempeño que evalúen el PC entre estudiantes universitarios, la calidad de las medidas obtenidas, que se presentarán en un artículo que está en preparación, confirman que con este tipo de pruebas se pueden obtener resultados de los niveles de desarrollo del PC alcanzados por los estudiantes universitarios con las cuales orientar las estrategias para el desarrollo de esta competencia en las universidades.

A continuación recogemos los pasos que seguimos en este proceso y señalamos para cada uno de ellos las principales lecciones aprendidas.

Comprensión del escenario base en su contexto original. Aunque el universo de la prueba puede construirse desde cero, partiendo de problemáticas completamente locales, partir de un escenario de base obliga a no perder de vista la interacción entre la prueba, su contexto y su entorno de aplicación. La ventaja, y el reto, de las pruebas basadas en el desempeño es que evocan una realidad compartida. La prueba ha de responder e interactuar con esta realidad y quienes crean una prueba deben ser conscientes de dicha interacción. El proceso debe comenzar con una conversación amplia sobre las múltiples maneras en las que una problemática es recibida, tramitada e interpretada por un grupo social.

Aplicación de un protocolo de pensamiento en voz alta. La aplicación de este protocolo provee la evidencia necesaria para validar la eficacia de la adaptación para representar y evaluar el constructo. En efecto, la visión del equipo evaluador con respecto a la relación

entre la prueba y su entorno debe corroborarse con la experiencia subjetiva de quienes responden la prueba y los criterios de evaluación deben confrontarse con lo que, en el caso en cuestión, resulte particularmente retador para ellos. El pensamiento en voz alta permite hacer esta verificación.

Aplicación piloto. La aplicación piloto implica realizar una convocatoria amplia, para contar con respuestas suficientes para la validación de la prueba. En nuestra implementación, contar con estudiantes de siete universidades colombianas, de facultades y programas diferentes y de características demográficas diversas, facilitó este proceso. En este ejercicio resulta crítico lograr motivar a los estudiantes participantes para que sus respuestas correspondan a las que darían en una aplicación real.

Entrenamiento de codificación y revisión. El proceso de codificación es arduo e implica la capacitación de un equipo que conozca la prueba a profundidad, además de participar en un proceso de evaluación colegiada, que garantice una alineación en las maneras de aproximarse a las pruebas y de interpretar los descriptores de la rúbrica. La codificación debe contar con un equipo de revisores que resuelvan las divergencias de dos maneras: comentando las codificaciones discrepantes con los codificadores involucrados e interviniendo directamente en la codificación. Es recomendable que, en una primera fase se le dé prioridad a la conversación con los codificadores discrepantes para construir con ellos una interpretación común de la rúbrica. La intervención directa del revisor se debe reservar solo para la fase final de la codificación.

Medición de confiabilidad. El coeficiente de *Generalizabilidad*, desarrollado como parte de la teoría que lleva su nombre (Shavelson & Webb, 1991), es la medida que permite determinar la confiabilidad de las evaluaciones de desempeño, en particular de su codificación. La verificación de que se alcanzan niveles satisfactorios en estas medidas es crucial para determinar el número de codificaciones que es preciso hacer de cada prueba y los niveles de discrepancia entre codificadores que se pueden tolerar.

Sincronicidad de la aplicación virtual. En esta oportunidad, la evaluación se realizó en entornos tanto virtuales sincrónicos como asincrónicos. Esto permitió adaptar la logística a los contextos particulares de las diferentes universidades. Sin embargo, es probable que los estudiantes sean más proclives a la dispersión en entornos virtuales asincrónicos, los cuales, como es natural, no están socialmente regulados. El análisis del efecto que esta variación pudo haber tenido en los resultados se presentarán en el artículo de análisis de resultados.

En suma, todas las fases del proyecto hasta aquí descritas contribuyeron para que se consolidara la prueba, en donde no solamente evaluamos las habilidades de PC de los estudiantes sino las de nosotros mismos como codificadores. Fueron necesarias discusiones, talleres y reuniones constantes para llegar a consensos que nos permitieran construir un instrumento evaluativo fiable y lo más objetivo posible. Posiblemente la mayor fortaleza del proyecto fue establecer canales de cooperación entre varias universidades, lo que

permitió contar con distintas percepciones respecto al desarrollo y evaluación de las competencias genéricas y del PC en Colombia, llevando a enriquecer el proceso de adaptación y validación de la prueba. Asimismo, dicha cooperación permitió obtener la participación de estudiantes de distintos contextos socioeconómicos en la validación y la aplicación final del instrumento.

Cabe también resaltar aquí que, los resultados de la prueba muestran que una competencia genérica como el PC puede evaluarse consistentemente como una habilidad transferible. La experiencia aporta evidencia para soportar la presuposición de transferibilidad que asumimos en esta investigación, lo cual abre las puertas para que podamos diagnosticar y promover el PC en estudiantes de diferentes áreas del conocimiento con el fin de enriquecer su proceso de formación. Esto muestra la importancia de la presencia del PC y su desarrollo en contextos de toma de decisión y resolución de problemas, por lo que resulta fundamental promover y crear métodos de evaluación que permitan fortalecer esta competencia en las instituciones de educación superior. Desde esta perspectiva, el proceso realizado en torno a la adaptación, validación y aplicación de la prueba de PC contribuye a avanzar en la creación de métodos evaluativos que permitan diagnosticar el desempeño en ésta y otras competencias en Colombia y América Latina.

REFERENCIAS

- Alto Comisionado de las Naciones Unidas para los Refugiados (ACNUR) (2022). *Refugiados y Migrantes*. Recuperado de: <https://refugeesmigrants.un.org/es/definitions>.
- Bonk, C. J., & Smith, G. S. (1998). "Alternative instructional strategies for creative and critical thinking in the accounting curriculum." *Journal of Accounting Education*, 16(2), 261–293.
- Braun, H.I.; Williamson, E.; Shim, C. & Borowiec, K. (2019) Legacy admissions. Unpublished Performance Assessment of Critical Thinking.
- Case, R. (2005). "Moving critical thinking to the main stage." *Education Canada*, 45(2), 45–49.
- Castells, L. (2007). Los protocolos de pensamiento en voz alta como instrumento para analizar el proceso de escritura. *RESLA 20*, 27–35. Recuperado de: <https://dialnet.unirioja.es/servlet/articulo?codigo=2514275>.
- Fischer, S. C.; Spiker, V. A. & Riedel, S. L. (2009). "Critical thinking training for army officers, volume 2: A model of critical thinking. (Technical Report)." Arlington, VA: *U.S. Army Research Institute for the Behavioral and Social Sciences*.
- Fries, L. (2019). Las mujeres migrantes en las legislaciones de América Latina: análisis del repositorio de normativas sobre migración internacional del Observatorio de Igualdad de Género de América Latina y el Caribe. Editorial CEPAL, 157(40), p. 84, Julio 2019. Recuperado el 23 de abril de 2020, de: <https://www.cepal.org/es/publicaciones/44655-mujeres-migrantes-legislaciones-america-latina-analisis-repositorio-normativas>. ISSN 1564-4170.
- Halpern, D. F. (1998). "Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring." *American Psychologist*, 53(4), 449–455.

- Harboe, F. (19/12/2018). *Inmigración en Chile, clave para el desarrollo*. Cooperativa, Opinión Política. Recuperado el 14 de mayo de 2020, de: <https://opinion.cooperativa.cl/opinion/politica/inmigracion-en-chile-clave-para-el-desarrollo/2018-12-19/063950.html>.
- Hyttinen, H. & Toom, A. (2019). Universidad de Helsinki. Prueba *Crisis de refugiados*.
- Kosslyn, S. B. & Nelson, B. (2017). *Building the intentional university. Minerva and the future of higher education*. Cambridge, MA: MIT Press.
- Lipman, M. (1988). "Critical thinking—What can it be?" *Educational Leadership*, 46(1), 38-43.
- Ministerio de Educación Nacional (MEN) (2006). Estándares Básicos de Competencias en Lenguaje, Matemáticas, Ciencias y Ciudadanas. ISBN 958-691-290-6. Recuperado el 25 de mayo de 2020, de: https://www.mineducacion.gov.co/1621/articles-340021_recurso_1.pdf.
- Naciones Unidas (2022a). *Migración*. Recuperado el 5 de mayo de 2020, de: <https://www.un.org/es/global-issues/migration>.
- Naciones Unidas (2022b). *Pacto Internacional de Derechos Civiles y Políticos*. Recuperado el 3 de mayo de 2020, de: <https://www.ohchr.org/es/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights>.
- Norris, S. P. (1989). "Can we test validly for critical thinking?" *Educational Researcher*, 18(9), 212-6.
- Paul, R. W. (1992). "Critical thinking: What, why, and how?" *New Directions for Community Colleges*, 1992(77), 3–24.
- Ramón, M. Á. & Guahichullca, L. (2018). La influencia negativa de la migración en el proceso educativo de la ciudad de Machala. *Revista Universidad y Sociedad*, 10(3), 217-225. Epub 02 de junio de 2018. Recuperado el 19 de abril de 2020, de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2218-36202018000300217&lng=es&tlng=es.
- Ronderos, N.; Shavelson, R. J.; Holtsch, D.; Zlatkin-Troitschanskaia, O. & Solano-Flores, G. (2021). Pruebas internacionales de desempeño para la evaluación del pensamiento crítico: marco para la traducción y adaptación. *Journal of Supranational Policies of Education*, (13), 62–87. <https://doi.org/10.15366/jospoe2021.13.003>.
- Rolon, N. (2014). Pensamiento crítico y docencia. Breves reflexiones de su aporte y riqueza. *Revista Didac*, (64), 18-23. http://revistas.iberomx.mx/didac/uploads/volumenes/18/pdf/Didac_64.pdf.
- Shavelson, R. J. (2013). "On an approach to testing and modeling competence." *Educational Psychologist*, 48(2), 73–86. doi:10.1080/00461520.2013.779483.
- Shavelson, R. J.; Zlatkin-Troitschanskaia, O.; Beck, K.; Schmidt, S. & Mariño, J. P. (2019) Assessment of University Students' Critical Thinking: Next Generation Performance Assessment, *International Journal of Testing*, 19:4, 337-362, DOI: 10.1080/15305058.2018.1543309.
- Shavelson, R. J.; Zlatkin-Troitschanskaia, O. & Mariño, J. P. (2018). International performance assessment of learning in higher education (iPAL): Research and development. En Troitschanskaia, O., Toepper, M., Pant, H.A., Lautenbach, C., Kuhn, C. (Eds.), *Assessment of learning outcomes in higher education* (1ra ed., pp. 193-214). EEUU: Springer.
- Shavelson, R.J. & Webb, N.M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: SAGE Publications.
- Schendel, R. & Tolmie, A. (2017). Beyond translation: adapting a performance-task-based assessment of critical thinking ability for use in Rwanda. *Assessment and Evaluation in Higher Education*, 42(5), 673–689. <https://doi.org/10.1080/02602938.2016.1177484>.

- Silva, E. (2008). "Measuring Skills for the 21st Century [Report]." Washington, D. C.: Education Sector. Retrieved from http://www.educationsector.org/usr_doc/MeasuringSkills.pdf.
- Wheeler, P. & Haertel, G. D. (1993). *Resource handbook on performance assessment and measurement: A tool for students, practitioners, and policymakers*. Berkeley, CA: Owl Press.
- White House (2020). Tomado de la página web: <https://www.whitehouse.gov/issues/immigration/%20%20---%20https://www.whitehouse.gov/briefings-statements/president-donald-j-trump-honoring-commitment-protect-american-workers-temporarily-pausing-immigration/>