

## Sistematización de pruebas de desempeño en pensamiento crítico: una experiencia de construcción interdisciplinar.

Systematization of performance tests in critical thinking: an interdisciplinary construction experience.

Maestra Nina Crump\*

Fil. Carolina Sepúlveda\*\*

Ing. Andrés Fajardo\*\*\*

Mag. Adriana Vera Aguilera\*\*\*\*

### Resumen

En este artículo se sistematiza la experiencia de creación de pruebas estandarizadas de pensamiento crítico (PC) y su evaluación, proyecto desarrollado por un grupo académico interdisciplinar. Se presentan las diferentes etapas del trabajo y se describe la metodología aplicada para concretar el modelo de evaluación. Se enfatiza en los aportes de los participantes para proponer una definición de PC y determinar los criterios para evaluar las habilidades asociadas a esta competencia. Por último, se exponen los aprendizajes que deja este ejercicio, con lo cual se busca contribuir a la puesta en marcha de otras iniciativas de evaluación del PC en la educación superior.

**Palabras clave:** pensamiento crítico, pedagogía, evaluación, interdisciplinariedad, pruebas estandarizadas.

### Abstract

In this article, the experience of creating standardized assessment of critical thinking and its evaluation is systematized, a project developed by an interdisciplinary academic group. The different stages of the work are presented and the methodology that is applied to specify the evaluation model is presented. In addition, emphasis is placed on the contributions of the participants to propose a definition of critical thinking and determine the criteria to evaluate

---

\* Francocolombiana; Abogada; Maestría en Educación; Universidad de los Andes, Bogotá, Colombia. Correo electrónico: [n.crump93@uniandes.edu.co](mailto:n.crump93@uniandes.edu.co).

\*\* Colombiana; Filósofa; Universidad de los Andes; Bogotá; Colombia. Correo electrónico: [carolinasepulveda@uniandes.edu.co](mailto:carolinasepulveda@uniandes.edu.co).

\*\*\* Colombiano; Ingeniero industrial; Universidad de los Andes; Bogotá; Colombia. Correo electrónico: [ad.fajardo10@uniandes.edu.co](mailto:ad.fajardo10@uniandes.edu.co).

\*\*\*\* Colombiana; Comunicadora social; Magíster en Estudios Culturales; Universidad de los Andes; Bogotá; Colombia. Correo electrónico: [ap.vera@unianes.edu.co](mailto:ap.vera@unianes.edu.co).



the skills in this competition. Likewise, reference is made to the lessons learned from this exercise, which is intended to contribute to the implementation of other critical thinking evaluation initiatives in higher education.

**Keywords:** critical thinking, pedagogy, assessment, interdisciplinarity, standardized tests.

## Introducción

El desarrollo del pensamiento crítico (PC) es uno de los grandes retos de la educación superior en el siglo XXI. Para que las universidades puedan avanzar con firmeza en el desarrollo de esta competencia, es necesario, en primer lugar, diseñar actividades en las que se lleven a la práctica habilidades como analizar fuentes de información, seleccionar información relevante, reconocer y evaluar posturas, examinar posibles consecuencias de distintas alternativas y emitir y sustentar juicios razonados, entre otras. En segundo lugar, es preciso configurar instrumentos de evaluación que den cuenta de los niveles alcanzados por los estudiantes en esta competencia. Solo mediante la implementación de instrumentos de este tipo es posible monitorear el desarrollo del PC en diferentes grupos de estudiantes, analizar resultados y elaborar conclusiones sobre las estrategias de su enseñanza.

En la Universidad de los Andes, de la ciudad de Bogotá, un grupo de profesores y estudiantes de maestría, conformado por equipos de las facultades de Ingeniería, Educación y Derecho, y los centros de Español y de Evaluación, ha trabajado durante los años 2017, 2018 y 2019 en la creación y validación de evaluaciones de desempeño que permitan medir habilidades de PC en estudiantes de educación superior. Este artículo tiene como objetivos principales: (1) describir e interpretar esta experiencia interdisciplinar, lo cual comprende un análisis de las consideraciones teóricas y los diálogos reflexivos que guiaron la ejecución del proyecto, así como la reconstrucción y el ordenamiento de los procesos que se llevaron a cabo en la elaboración y validación de los instrumentos de evaluación, y (2) divulgar los aprendizajes alcanzados sobre el diseño de las pruebas y su correlación con las habilidades del PC que se buscaron evaluar, con el fin de que puedan ser examinados por investigadores de prácticas similares.

Para empezar, en el artículo se hace una descripción del contexto en el que surge esta iniciativa: el marco internacional en el que se inscribe la actividad desarrollada por los equipos de la Universidad de los Andes (sección 1). Después, se explica la metodología que se aplicó para cumplir con el objetivo de diseñar y validar las distintas pruebas de PC y sus respectivas rúbricas de evaluación (sección 2). Luego, se aborda el ejercicio conjunto de reflexión que le permitió al grupo interdisciplinar

construir una definición de PC y determinar las habilidades que se consideran de interés para su evaluación, así como establecer las condiciones comunes de las pruebas. En este apartado se exponen discusiones sobre el marco teórico inicial que orientó la propuesta y la manera cómo, a partir de estas, se fijaron las directrices para crear los instrumentos de evaluación (secciones 3, 4 y 5). Posteriormente, se profundiza en los procedimientos que se llevaron a cabo para validar las pruebas y ensayar las rúbricas de evaluación (secciones 6 y 7). Finalmente, a manera de conclusiones, se exponen algunas de las lecciones aprendidas durante cada etapa.

## 1. Contextualización de la iniciativa del diseño de pruebas de pensamiento crítico

El PC hace parte de las competencias genéricas<sup>1</sup>, es decir, aquellas que son fundamentales para el desarrollo social y laboral de un individuo, así como para su participación ciudadana: son aquellas que deben desarrollar todas las personas, independientemente de su formación, pues son indispensables para el desempeño académico y laboral (Icfes, 2015). En estas competencias se destaca la importancia de integrar, sintetizar y aplicar conocimientos para resolver problemas que se presentan en la cotidianidad; su desarrollo no se centra en la adquisición o la extensión de contenidos de conocimientos específicos que haya adquirido un individuo, sino en cómo usa las habilidades que tiene para resolver retos de la vida diaria (Herde, Wüstenberg & Greiff, 2016).

En la actualidad, las instituciones de educación superior dedican esfuerzos importantes para fortalecer en sus estudiantes el desarrollo del PC. Sin embargo, no se dispone de mecanismos de evaluación que permitan dar cuenta clara de los resultados que estos esfuerzos producen (Hatcher, 2011). En consecuencia, resulta necesario diseñar evaluaciones e instrumentos que permitan obtener medidas válidas del desarrollo de esta competencia en programas de educación superior (Zlatkin-Troitschanskaia, Pant & Coates, 2016), tanto para producir diagnósticos preliminares de los niveles de PC que alcanzan los estudiantes de pregrado, como para monitorear su avance en el transcurso de los estudios universitarios.

---

<sup>1</sup> En Colombia, con el propósito de proporcionar información sobre el grado de desarrollo de estas competencias, el Examen de Estado de Calidad de la Educación Superior (Saber Pro) contiene una sección de competencias genéricas que comprende cinco módulos: 1) Lectura crítica, 2) Razonamiento cuantitativo, 3) Competencias ciudadanas, 4) Comunicación escrita e 5) Inglés (Icfes, 2018).



Para responder a estas necesidades, bajo el liderazgo del profesor Richard Shavelson, se puso en marcha en el 2015 el proyecto International Performance Assessment of Learning (iPAL), el cual involucra a académicos e investigadores de varios países (Estados Unidos, Alemania, Canadá, México, Suiza, Israel, Reino Unido, Finlandia y China). Su principal propósito es promover un trabajo conjunto de desarrollo de pruebas estandarizadas de desempeño para la medición de PC, actividad que se enfoca en las instituciones de educación superior (Centro de Evaluación Educativa, 2017). En resumen, iPAL tiene como objetivo.

[...] construir evaluaciones en las que los evaluados se enfrentan a situaciones de alta complejidad. Las especificaciones del diseño de las situaciones problemáticas exigen a los evaluados evidenciar, en el planteamiento de una solución, el conjunto de condiciones constituyentes del PC. Los protocolos de la evaluación y, especialmente, los controles en la aplicación de las rúbricas de calificación aseguran la estandarización de las medidas que se producen (Centro de Evaluación Educativa, 2017: 1).

Dados los intereses del grupo interdisciplinar de la Universidad de los Andes en el desarrollo del PC, en el año 2017 se origina el proyecto PALSA (Performance Assessment of Learning at Andes) como un subproyecto de iPAL. Además de diseñar pruebas de PC, PALSA busca obtener información sobre los niveles de PC alcanzados por los estudiantes de la Universidad, lo cual resulta útil para adelantar estudios comparativos nacionales e internacionales. Por otra parte, hay un interés especial en analizar el potencial pedagógico de las pruebas en la formación del PC. Es decir, su posible impacto en las metodologías y prácticas pedagógicas de los profesores participantes. En particular, se espera que los aprendizajes y conocimientos derivados de esta experiencia interdisciplinar enriquezcan los programas de enseñanza e incidan en el mejoramiento de competencias.

## Involucrados

Como se mencionó, el proyecto PALSA se desarrolla en el contexto específico de la Universidad de los Andes: es una iniciativa de un grupo interdisciplinar de profesores y evalúa a estudiantes de distintos programas de la Universidad. Cada grupo estuvo representado por profesores asociados o directores de área, quienes se desempeñaron como coinvestigadores. Además, el proyecto contó con dos asistentes de investigación, que acompañaron los procesos de las unidades académicas, y con la asesoría externa del profesor Richard Shavelson. A continuación, se describen las razones que motivaron a los distintos equipos a integrarse al proyecto PALSA.

### Grupo de Ingeniería Industrial

El programa de pregrado de Ingeniería Industrial incluye en su plan de estudios la asignatura Dinámica de sistemas, la cual cursan estudiantes de cuarto y quinto semestre. Uno de sus módulos se enfoca en el enriquecimiento de las perspectivas para el diseño y la construcción de modelos sistémicos aplicando la heurística crítica de sistemas de Ulrich (2005). El grupo de ingeniería está conformado por los profesores de este curso: Camilo Olaya, coordinador del curso y director del Departamento de Ingeniería industrial; Laura Guzmán y Juliana Gómez, profesoras de cátedra.

Para el grupo es fundamental que sus estudiantes desarrollen habilidades de PC que les permitan abordar de forma integral situaciones a las que se enfrentarán en su vida laboral y profesional; por ejemplo, intervenir un sistema, lo cual implica, entre otras acciones, analizar y cuestionar información, considerar las perspectivas de los involucrados, tomar una postura frente a las problemáticas que se esbozan, trazar una frontera de alcance y entender las consecuencias de la aplicación de un modelo.

### Grupo de Educación

La Facultad de Educación tiene como uno de sus objetivos principales investigar sobre la pedagogía en Colombia y sus transformaciones. Su curso El sentido de la educación, dirigido a estudiantes de la Maestría en Educación, se centra en las pedagogías críticas y el desarrollo del PC, lo cual llevó al profesor Andrés Mejía a vincularse al proyecto PALSA. Mejía (2002), desde sus estudios de doctorado en la Universidad de Hull (Reino Unido), se ha interesado en observar y evaluar diferentes aspectos del PC, lo cual recoge en algunos apartados de su tesis titulada *A critical systemic framework for studying knowledge imposition in pedagogy*. Adicionalmente, Mejía ha trabajado con diversas universidades en la articulación de la formación ciudadana y el PC. Su reflexión y producción académica contribuyeron a enriquecer las discusiones del grupo. En especial, su artículo, en coautoría con Cascante (2007), *Hacia una dinámica de sistemas crítica: un marco conceptual para investigación y una ilustración en educación* guió la construcción de una definición inicial de PC.

### Grupo del Centro de Español

El Centro de Español (CE) es una unidad adscrita a la Vicerrectoría de Asuntos Académicos de la Universidad de los Andes. Dentro de sus funciones está desarrollar, fortalecer y acompañar el dominio académico del lenguaje de los estudiantes de pregrado de la Universidad. Para ello, se vale de estrategias como



los cursos con alto componente de escritura (cursos tipo E), los cursos básicos de Español y las tutorías de escritura, lectura y oralidad.

Su enfoque pedagógico comprende tres competencias: i) analítica, que se vincula con el manejo de la información; ii) reflexiva, relacionada con la valoración de la información y iii) argumentativa, que implica la toma de posición razonada frente a una situación problemática (Escallón y Forero, 2016). Para el CE, y de acuerdo con su misión institucional, estas tres competencias contribuyen al desarrollo del PC de manera transversal. Atendiendo a esta consideración, los profesores Andrés Forero, Tatiana Cáceres, David Alvarado y Andrés Bautista se integraron al proyecto PALSA.

### Grupo de Derecho

El grupo de Derecho se vinculó al proyecto PALSA debido al interés específico de una de sus docentes, Betsy Perafán, por plantear actividades y formas de evaluación que debatan lógicas tradicionales del ejercicio del derecho.

Los intereses de esta área en relación con el PC parten de la asignatura Introducción al Derecho, a la que asisten estudiantes de primer semestre del programa de Derecho y de otros, como Economía y Ciencia Política. El contenido de este curso se relaciona con lo que se conoce como Aprendizaje Basado en Problemas (ABP). Para el grupo de Derecho es importante que el estudiante se enfrente a situaciones problemáticas de la vida real; en ese sentido, en el marco de la medición del PC, se pretende evaluar la forma en la que el estudiante identifica y rastrea diferentes variables de un hecho, cuestiona creencias y perspectivas y construye soluciones. Adicionalmente, cabe mencionar que el interés de la profesora Perafán en el desarrollo del PC se basa en cuestionar la idea generalizada del derecho de proponer soluciones polarizadas, en las que se determina que solamente una de las partes involucrada en una problemática está en lo correcto y la otra debe compensarla. En otras palabras, con el desarrollo de ejercicios de PC en esta clase, se busca que los estudiantes reflexionen acerca de soluciones legales que no entren en la lógica de víctima y victimario, sino que contemplen alternativas benéficas para la sociedad y, particularmente, para las dos partes involucradas.

### Grupo Centro de Evaluación

El grupo del Centro de Evaluación Educativa (CEE) que ha trabajado en el proyecto, además de los autores de este artículo (Nina Crump, Carolina Sepúlveda, Andrés Fajardo y Adriana Vera), cuenta con la participación de Julián Mariño y Adriana Molina. El CEE, unidad adscrita a la Facultad de Educación de la Universidad de los

Andes, desarrolla proyectos de medición y evaluación educativa en todos los niveles de formación<sup>2</sup> para establecer aspectos como el grado de competencias alcanzado por los estudiantes en distintas áreas de conocimiento. A partir de la experiencia de trabajo con el proyecto internacional iPAL, el CEE diseñó una propuesta de investigación que se consolidó como el proyecto PALsA, al que se integraron las áreas mencionadas.

El Centro de Evaluación se ha interesado en incluir el componente de PC en varios de los proyectos que desarrolla; por ejemplo, en aquellos que se adelantan en colegios. En este sentido, el CEE busca que algunos de los hallazgos derivados del diseño y evaluación de pruebas de desempeño de PC en educación superior puedan aplicarse a los niveles de educación secundaria y media.

## 2. Metodología

La metodología para el diseño de las pruebas de desempeño de PC y la rúbrica de evaluación comprendió dos fases: i) fase de desarrollo y ii) fase de aplicación. En ambas fases se consideró la implementación de distintos tipos de actividades e instrumentos de recolección de información enmarcadas en el enfoque cualitativo. En este artículo, solo se describirá la primera fase, ya que la fase de aplicación aún se encuentra en marcha.

### Fase de desarrollo

- a) Revisión de literatura especializada y construcción del marco teórico. En esta etapa, desarrollada especialmente en el primer ciclo del proyecto, se construyeron los lineamientos teóricos en los que se fundamentan las pruebas. En este punto, el grupo propició discusiones respecto a los componentes que se consideran relevantes en los ejercicios de PC y su respectiva evaluación. A partir de la información consultada se propuso una definición de PC y se concretaron acuerdos sobre las condiciones de las pruebas. Además, se determinaron las habilidades que se evaluarían, lo que dio paso a esquematizar una primera versión de rúbrica.
- b) Talleres de construcción de pruebas de desempeño de PC. Se propusieron como espacios en los que se trabajara de manera colectiva en el

---

<sup>2</sup> En Colombia, el sistema escolar comprende: la educación preescolar, la educación básica (primaria, cinco grados, y secundaria, cuatro grados), la educación media (dos grados) y la educación superior (Ministerio de Educación Nacional, 2010).



planteamiento de los elementos constituyentes de las pruebas (situaciones problemáticas, tareas y documentos o fuentes para apoyar su desarrollo, entre otros). En los talleres, los avances de las pruebas se sometieron a revisión de los pares del proyecto, es decir, los grupos disciplinares de cada unidad académica.

- c) Elaboración de las pruebas de desempeño y de la rúbrica. En esta etapa, cada grupo elaboró distintas versiones de sus pruebas. Para ello, se tuvo en cuenta el marco teórico, así como el estado del arte construido y la interpretación de ejemplos de pruebas de desempeño desarrolladas a nivel internacional, como la Evaluación de resultados de aprendizaje de la educación superior (AHELO por sus siglas en inglés) y el Collegiate Learning Assessment (CLA) (Shavelson, Troitschanskaia & Mariño, 2017). Las pruebas diseñadas pasaron por la revisión de los pares del proyecto y se construyeron y modificaron a la luz de las observaciones y recomendaciones de expertos académicos como Richard Shavelson. Posteriormente, se precisaron los criterios de evaluación y los niveles de desempeño de la rúbrica, de acuerdo con las habilidades del PC de interés.
- d) Validación de las pruebas. En este ciclo, las pruebas se aplicaron a una pequeña población de estudiantes con el fin de recoger información respecto a su funcionamiento (claridad de las instrucciones, pertinencia disciplinar, grado de dificultad, valoración de las fuentes y evaluación de las propuestas de solución de los participantes). En este sentido, se aplicaron instrumentos de recolección de información como las entrevistas semiestructuradas individuales y un protocolo con preguntas abiertas y cerradas. Estas herramientas se configuraron en el marco de dos ejercicios, los cuales se efectuaron en el orden que se presenta a continuación.
  - ) Pensamiento en voz alta. Instrumento en el que el evaluado, individualmente, narra las ideas y el hilo de pensamiento que elabora a medida que resuelve la prueba (Castells, 2007). Los evaluados fueron estudiantes que aprobaron con buenos resultados la asignatura de cada programa en el que se aplicarían las pruebas, monitores o asistentes graduados de esas asignaturas. En total, este ejercicio lo realizaron 17 participantes: cuatro (4) en la prueba de Ingeniería Biomédica (Centro de Español), cinco (5) en la prueba de Derecho, cuatro (4) en la prueba de Ingeniería Industrial y cuatro (4) en la prueba de Educación. La información fue recogida mediante un protocolo de pensamiento en voz alta por medio de preguntas abiertas y cerradas en entrevistas semiestructuradas.



- J) Pilotaje. Implementación de la prueba, en una misma sesión, a un grupo de estudiantes con un perfil similar al de quienes presentarán el ejercicio final. En este pilotaje se aplicaron 145 pruebas (145 participantes), distribuidas de la siguiente manera: 16 pruebas de Ingeniería Industrial, 53 pruebas de comunicación de Ingeniería Biomédica (prueba del Centro de Español), 46 pruebas de Educación y 30 pruebas de Derecho. Cada prueba fue evaluada por dos pares del equipo del proyecto. Las calificaciones fueron contrastadas para identificar acuerdos y discrepancias, de tal manera que se determinaron los criterios con los resultados más bajos y más altos, lo cual condujo a un ajuste de las descripciones de los criterios y los niveles de desempeño de la rúbrica.

Dada la importancia de la validación de los instrumentos de evaluación, y su impacto en su rediseño, en las secciones 6 y 7 del artículo, se explicará con más detalle cómo se aplicaron las herramientas de pensamiento en voz alta y pilotaje, así como los resultados obtenidos. De esta manera, se hace evidente que el proceso de creación de las pruebas y las rúbricas es iterativo, pues exige revisiones, ensayos, ajustes y reinterpretaciones.

### 3. Discusiones

En este apartado se presentan, de forma sucinta, los principales debates entre los participantes del proyecto para concretar los instrumentos más idóneos en la evaluación del PC. Se partió de la necesidad de proponer una definición de PC y especificar los desempeños que permiten evidenciar su desarrollo. La construcción de una forma común de entender el PC se dio paulatina y progresivamente durante distintos encuentros en los que se examinaron las perspectivas de las unidades académicas en torno a esta competencia. Si bien durante las discusiones los equipos se esforzaron por llegar a acuerdos, también cada uno de ellos se interesó por conservar aquellos aspectos del PC que vincula con su marco disciplinar.

A continuación, se presentan siete asuntos que fortalecieron el debate: i) la definición de PC; ii) el carácter genérico o específico del PC y las posibilidades de elaborar pruebas transferibles; iii) la disposición o el componente actitudinal del estudiante; iv) los componentes del contexto o situación de las pruebas; v) el enfoque basado en el multiperspectivismo; vi) el componente ético; vii) el componente de escritura.



### 3.1 Una definición de pensamiento crítico

Existen diferentes aproximaciones a lo que significa el PC y su medición. En algunas, es común que se integren competencias que se evidencian en la toma de posición razonada frente a una situación problemática o en la resolución de un problema en el que es necesario examinar un contexto, en el cual divergen y se complementan distintas perspectivas (Ennis, 2018).

En el marco del proyecto iPAL se hace referencia al PC como el “proceso de conceptualizar, analizar, sintetizar y evaluar información para resolver un problema, decidir sobre un curso de acción, encontrar una respuesta a una pregunta o llegar a una conclusión” (Shavelson, Troitschanskaia & Mariño, 2018). A partir del análisis de esta caracterización del PC, y atendiendo a las inquietudes, experiencias y conocimientos de cada equipo, en la iniciativa PALSA se definió el PC como: la capacidad que le permite a una persona proponer una solución o adoptar conscientemente una posición frente a una situación problemática, después de comprender los hechos, cuestionar sistemáticamente las perspectivas, los intereses y los argumentos de las partes involucradas en dicha situación, y evaluar las consecuencias o repercusiones de la decisión por la que se optó. En este proceso de resolución, se valora, particularmente, el curso de acción planteado y la coherencia con la que este se explica y justifica. Esta aproximación al PC que propone PALSA orientó la elaboración de los instrumentos de evaluación; en otras palabras, configura un modo en que se considera que puede desarrollarse, evidenciarse y evaluarse el PC.

En el proceso de definición del PC fueron surgiendo criterios evaluativos y conceptuales anclados en los intereses disciplinares de cada unidad. Es decir, desde su experticia, como se evidencia en las pruebas diseñadas, cada área ponderó los aspectos del PC que considera relevantes para su evaluación. Por ejemplo, para el grupo de Ingeniería Industrial la teoría del multiperspectivismo de Ulrich (2005) resulta de mayor importancia que para otras áreas. Para el caso del grupo de Derecho, la prueba, además de incluir el análisis de perspectivas, se enmarca en un ejercicio pedagógico más amplio, pues el aprendizaje en el aula y el componente ético se consideran fundamentales para su resolución. Por su parte, el enfoque del Centro de Español ubica a la calidad de la escritura como piedra angular de la medición del PC. Por último, para el grupo de Educación también es preponderante el análisis de las múltiples perspectivas que se involucran en una situación problemática.

### 3.2 Carácter genérico o específico del PC y posibilidades de elaborar pruebas transferibles

Otro de los debates a los que se enfrentó el grupo PALSA, y para el cual todavía no hay una respuesta definitiva, corresponde a si el PC es una competencia genérica o si está condicionada por conocimientos específicos. En la literatura relativa a PC, algunos argumentan que este es un dispositivo independiente de conocimientos particulares, mientras que otros sostienen que se asocia con la adquisición de conocimientos disciplinares (Moore, 2011). Si bien en PALSA se pensó desde el principio en diseñar pruebas propias para cada área, también se consideró deseable que estudiantes de distintos programas académicos puedan darles respuesta.

En términos generales, permanece la inquietud sobre la posibilidad de que los estudiantes con conocimientos específicos de un área alcancen un mejor desempeño con respecto a aquellos que no los tienen. De hecho, luego de las aplicaciones finales, este es uno de los aspectos que el grupo quiere analizar para dar cuenta de la correlación entre los conocimientos de la disciplina, con la cual se asocia una prueba, y los resultados de los estudiantes.

En el caso de PALSA, se identifica, en un primer momento, que la prueba de Ingeniería Industrial requiere de habilidades particulares. Concretamente, exige la interpretación de la representación gráfica de un modelo sistémico y una propuesta de solución en el que ese modelo se intervenga. El uso de este tipo de representaciones es parte de los contenidos del curso Dinámica de sistemas; sin embargo, se considera que estos aspectos no constituyen una limitación de fondo puesto que, si la prueba se quisiera aplicar a estudiantes que no han tomado esta asignatura, la representación gráfica se podría reemplazar por un texto continuo (compuestos secuencialmente a partir de párrafos), acompañado de recursos gráficos más sencillos, y solicitar a los estudiantes presentar su respuesta en un texto argumentativo. Igualmente, aunque las pruebas de Ingeniería Biomédica y Derecho se relacionan con temas propios de cada área, biopsias percutáneas y denuncias en redes sociales, estas no demandan conocimientos particulares para proponer respuestas consideradas como buenas. Por su parte, la prueba de Educación también se diseñó de manera que pueda ser resuelta por estudiantes de diferentes disciplinas. Sin embargo, consta de fuentes de información en las que se presentan tendencias y teorías pedagógicas que pueden resultar exigentes para estudiantes de áreas distintas al programa de Educación.

Atendiendo a esta situación, el grupo PALSA recopiló información y evidencias para enriquecer el debate sobre la relevancia o no de los conocimientos específicos para el despliegue del PC. En principio, se optó por hablar de habilidades genéricas que



deberían exigirse a quienes acceden a la educación superior independientemente de su disciplina académica. Aunque las pruebas planteen situaciones problemáticas particulares, se espera que midan habilidades comunes del PC. En este sentido, se determinó que la situación problemática, a la que se enfrenta al estudiante para que tome una decisión o proponga un curso de acción, no se limite a recrear aspectos tan locales, sino que, por el contrario, se caracterice de una manera más estandarizada, lo que haría que las pruebas puedan ser comprendidas en distintos escenarios y presentadas por diversos públicos, cuestión que se adecúa a los requerimientos de iPALS respecto a la traducción, adaptación y validación de pruebas en diferentes contextos (Shavelson, Troitschanskaia & Mariño, 2018).

Ahora bien, en el grupo aún persiste el interrogante de hasta qué punto es posible que las pruebas sean transferibles; es decir, que se puedan medir de manera efectiva las mismas habilidades de PC en distintos contextos disciplinares, nacionales e internacionales. Además, ha surgido la pregunta de si son necesarios ciertos requisitos para poder presentar cada prueba (por ejemplo, que se trate de un estudiante de x o y semestre y que haya cursado ciertas asignaturas) y, por ende, si el conocimiento de un área otorga ventajas para la resolución de la prueba. Estos problemas, además de abordarse en las discusiones del grupo, se han explorado en los ejercicios de pensamiento en voz alta y en la aplicación de los primeros pilotos. Para analizar estas cuestiones, en PALSA se optó por:

- )] Elaborar una rúbrica base con criterios estándar que cada grupo debe, como mínimo, tener en cuenta para la evaluación de su prueba.
- )] Reconocer y señalar los aspectos que adquieren mayor relevancia en cada prueba sin dejar de lado que se mantengan dimensiones transversales.
- )] Comparar las respuestas dadas a una misma prueba por estudiantes de distintos niveles y programas académicos.

### 3.3 Disposición

En el grupo también se discutió sobre la disposición, es decir, el componente actitudinal que incentiva al estudiante a desplegar su PC. Esta consideración supone que un ejercicio de PC no solo comprende habilidades cognitivas, sino que incluye aspectos anímicos o emocionales que configuran motivaciones para enfrentarlo (Butler, 2012).

En un primer momento, las discusiones sobre la disposición se orientaron a establecer si el componente actitudinal incide en las habilidades asociadas al PC. En

las pruebas se enfrenta al estudiante a una situación problemática verosímil en la que convergen y contrastan diferentes perspectivas. Para activar el PC, el grupo PALSA estimó necesario detenerse en la construcción de esa situación. Así, se optó por recrear situaciones actuales, polémicas y cercanas a los estudiantes, pues se considera que estas favorecen la motivación, el interés y la disposición necesaria para resolver las pruebas.

Posteriormente, se dio paso a examinar si algún aspecto motivacional o actitudinal puede evaluarse en las pruebas. Al respecto, se reconoce la dificultad de medir estos aspectos en el texto final que, como respuesta, elabora el estudiante. Entonces, se propone considerar el proceso de configuración de la respuesta como el espacio en el cual se puede indagar por el componente actitudinal y su posible influencia en el desarrollo del PC.

Con la finalidad de recopilar información sobre este asunto, se aprovechó el espacio de los ejercicios de Pensamiento en voz alta. En las pruebas de PC se le pide al estudiante proceder de forma consciente y reflexiva, es decir, apropiándose de su proceso de comprensión, reflexión y toma de decisiones, lo cual puede rastrearse en los ejercicios de Pensamiento en voz alta. En estos, los estudiantes, a medida que acceden a la prueba (la situación problemática, el rol asignado, la tarea y las fuentes), informan sobre diferentes aspectos como: procesos cognitivos, inquietudes, impresiones sobre el material, valoración de las fuentes, influencia de creencias y conocimientos previos, configuración de su respuesta y grado de dificultad de la prueba, entre otros. En concreto, esta actividad sirvió para evaluar si las pruebas eran comprensibles y realizables, y si generaban interés.

### 3.4 Contexto de las pruebas

El diseño de evaluaciones también dio lugar a la discusión sobre los componentes que conforman el contexto de las pruebas de PC. Su integración adecuada permite proponer un ejercicio más sólido y llamativo para los estudiantes. En este orden de ideas, se determinó que el contexto debe comprender posibles escenarios a los que puede enfrentarse un estudiante en su actividad académica actual y profesional inmediata. Los componentes del contexto que se consideraron son:

- a) El rol asignado al estudiante. En términos temporales, debe ser próximo a la situación actual del estudiante. En este sentido, por ejemplo, pedirle que tome una decisión como presidente de una nación o una multinacional puede considerarse poco verosímil. Por el contrario, si al estudiante se le asigna un rol como asistente o analista, le será más sencillo ubicarse en la situación



problemática y determinar cuáles son las herramientas adecuadas para enfrentar dicha situación.

- b) La situación problemática. Debe ser verosímil y de fácil acceso, por ejemplo, puede hacer referencia a una situación expuesta en los medios de comunicación e inscribirse a un determinado ámbito: político, deportivo, científico, económico, ambiental, cultural, etc. Por otra parte, esta situación debe exponer un conflicto que requiere una comprensión completa y la revisión de fuentes para sugerir una solución o curso de acción, lo cual implica reconocer y analizar las perspectivas y los propósitos de los involucrados.

Una de las discusiones principales sobre este punto, se refiere a la pertinencia o no de reproducir fielmente una situación de la vida real (por ejemplo, conservando los perfiles de los involucrados o haciendo referencia al espacio y la temporalidad real de los sucesos). Por una parte, al presentar la situación problemática de esta manera, se estima que se genera familiaridad con los hechos y aumenta el interés por dar una respuesta. Por otra, se considera que proponer una situación tal como ocurre, u ocurrió en la realidad, puede condicionar la respuesta del estudiante, ya que no es claro hasta qué punto se enfrente a la situación con objetividad, es decir, sin basarse en juicios y razonamientos previos sobre lo planteado. Esta sigue siendo una discusión abierta para el grupo.

- c) El destinatario/audiencia (a quien se le propone un curso de acción). Un actor clave del contexto es el destinatario de la respuesta. Su consideración es de gran importancia porque da lugar al cuestionamiento de sus requerimientos e intereses. Dado que se espera que el estudiante realice un ejercicio consciente, sin la presión de responder a un único actor, con unos intereses marcados, se decidió que dentro de las instrucciones se aclare que los resultados serán expuestos ante un grupo diverso de personas que atienden a diferentes propósitos. De esta forma, se espera que no haya un sesgo frente a la respuesta esperada. Al sugerir un grupo heterogéneo como audiencia, se promueve que se consideren las múltiples perspectivas de los involucrados, lo cual exige un análisis más amplio de la situación. Además, esto estimula al estudiante a ser crítico frente al objetivo planteado por el destinatario.

### 3.5 Multiperspectivismo

Para el grupo PALSA una de las habilidades clave del PC es el análisis de las múltiples perspectivas que subyacen a una situación problemática. En el desarrollo

de las pruebas, se espera que los estudiantes sean capaces de comprender diferentes puntos de vista y determinar su convergencia, divergencia o complementariedad. Así mismo, se espera que asuman una actitud dialógica que propicie la discusión y el debate entre los puntos de vista de los involucrados en la situación y los propios.

Este enfoque se basa en los estudios del profesor Werner Ulrich sobre la heurística de sistemas sociales y la herramienta de crítica de frontera, expuestos en distintos artículos como A Brief Introduction to Critical Systems Heuristics (2005), Critical Systems Heuristics (CSH) (2010) y Critical heuristics of social systems design (1987). De acuerdo con Ulrich, los juicios de frontera posibilitan pensar de forma más completa una situación problemática. Para alcanzar este propósito, el analista de la situación debe preguntarse acerca de cuestiones como: ¿cuál es el objetivo del análisis?, ¿quiénes están involucrados en la situación problemática?, ¿quiénes son actores y quiénes, los expertos?, ¿cuál es su visión de los hechos?, ¿cuáles son sus intereses y motivaciones?, ¿cuáles son las consecuencias de la solución propuesta?, ¿quiénes son los afectados? (Ulrich, 1987).

En este orden de ideas, en las pruebas de PC diseñadas por el grupo PALSA, el análisis de perspectivas implica examinar, entre otras cuestiones, el contexto (político, social, cultural, científico, ético, etc.) de la situación problemática, considerar los hechos, identificar los actores y los expertos, así como sus posiciones, las razones que los llevaron a adoptarlas y los argumentos que aportan para sustentadas. De esta manera, en este ejercicio resulta fundamental desplegar las perspectivas para develar fines particulares, intereses, motivaciones, impresiones e ideales de los involucrados y las diferentes cuestiones que quedan al margen de sus consideraciones.

Finalmente, un análisis de perspectivas resulta fundamental para evaluar las consecuencias y repercusiones de una decisión, pues considera a los directos involucrados y los posibles afectados. En otras palabras, la evaluación de perspectivas permite que se conciba una solución que cobije a los diferentes actores y se examine si hay espacio para reconciliar las diferentes posiciones. También, promueve una apertura a propuestas alternativas y al entendimiento de los razonamientos que las justifican: de alguna manera, el análisis de perspectivas busca un tipo de entendimiento mutuo o conciliación, así se le da vital importancia a la comunicación racional y argumentada.



### 3.6 Ética

Otra de las discusiones fue en torno a la relevancia o no de suscitar reflexiones éticas que incidan en las respuestas de los estudiantes (Mejía, 2009). Al respecto, surgieron las siguientes preguntas.

- a) ¿Cómo medir el componente ético en una prueba de PC? Las partes acordaron la dificultad que trae la medición de este componente. Si bien la academia se ha interesado en su medición, se concluyó que las pruebas están enfocadas en la evaluación de habilidades cognitivas de PC.
- b) ¿La ética y el PC están relacionados? Las partes estuvieron de acuerdo en que una persona puede pensar críticamente y proponer una respuesta o curso de acción que se aleje o no contemple principios éticos, y viceversa. Esto permitió ratificar la decisión de no darle peso a este componente dentro de la medición del PC.
- c) ¿Cómo evitar que las posiciones éticas dificulten un análisis completo del PC? En este ejercicio, se considera posible que un estudiante pueda o no dar del todo una respuesta completa y robusta en términos de PC al involucrar sus principios éticos. En este sentido, la recomendación es que se evite plantear una situación problemática en la que prevalezcan los dilemas éticos.

### 3.7 Componente de escritura

El grupo debatió sobre la importancia del proceso de escritura en la configuración de una respuesta. Se acordó que la evaluación de este componente no se limitara a la revisión gramatical y ortográfica de los textos que produzcan los estudiantes, pues involucra, entre otros aspectos, la capacidad de reconocer una audiencia y, en esa medida, hacer comprensible un mensaje. Para el grupo PALSA, el ejercicio de escritura no es una actividad mecánica, pues implica, entre otras cuestiones, el análisis de la información, la valoración y ponderación de posiciones, un ejercicio argumentativo y un conocimiento retórico. En este orden de ideas, el grupo del Centro de Español propuso hablar de comunicación y no solo de escritura, por ello se acordó establecer el criterio de evaluación Escritura y comunicación eficaz.



#### 4. Diseño de las pruebas

Luego de estudiar el marco iPAL para el diseño de evaluaciones y discutir sobre las condiciones de las pruebas, las unidades académicas de PALSA se centraron en su elaboración. Para empezar, se determinó una estructura común y se especificaron sus principales características. Se trata de pruebas:

- ) De desempeño y estandarizadas.
- ) Mixtas (con una pregunta abierta y preguntas de respuesta múltiple).
- ) Basadas en un caso que asemeja una situación problemática de la vida real.
- ) Con una pluralidad de fuentes para examinar y evaluar.
- ) Donde se le asigna un rol específico al estudiante.
- ) En las que se debe proponer un curso de acción como respuesta.

Como se mencionó anteriormente, la evaluación de las habilidades de PC presenta diversos desafíos. Experiencias anteriores han evidenciado dificultades asociadas con la confiabilidad y la validez de las medidas. En este sentido, las especificidades de las pruebas y su estandarización responden a la necesidad de desarrollar exámenes homogéneos y fiables, que permitan hacer juicios válidos sobre los desempeños de los evaluados, así como comparar resultados entre distintos grupos y a través del tiempo (por ejemplo, entre los resultados de estudiantes de diferentes semestres de una misma disciplina).

Las pruebas constan de una pregunta abierta y algunas preguntas de respuesta múltiple y respuesta corta, pues se considera que una prueba mixta permite aislar y diferenciar las habilidades que se quieren medir. Primero, a partir de un caso que recrea una situación problemática y posible en la vida real, se plantea una pregunta abierta, interrogante central de la prueba, para que el estudiante, asumiendo un rol, idee una respuesta (de una página aproximadamente), específicamente, un curso de acción o propuesta de solución.

Después se formulan preguntas de respuesta múltiple para evaluar la credibilidad, la validez y el sesgo de las fuentes proporcionadas para la resolución del caso. Finalmente, a través de preguntas de respuesta corta, se indaga acerca de la necesidad de información adicional (a la cual no se refieren las fuentes y que podría ser útil para la resolución del caso), si se requieren conocimientos específicos para



responder la prueba y si el estudiante cambiaría su respuesta en caso de que el rol asignado fuera otro.

En este ejercicio se proponen diversas fuentes en distintos formatos, como tablas, gráficos y textos escritos, para que los estudiantes puedan analizar información que los aproxime a los hechos, los actores involucrados en la problemática, los expertos que la interpretan y las perspectivas que adoptan. De tal manera que puedan comprender la situación de forma más amplia para concretar un curso de acción.

## Descripción de las pruebas

### Centro de Español

La prueba del CE está dirigida a estudiantes de últimos semestres de la carrera de Ingeniería Biomédica<sup>3</sup> del curso Diseño de proyectos 2. En este ejercicio se busca que el estudiante tome una decisión en relación con una situación problemática y a la que un ingeniero biomédico se podría ver enfrentado en la cotidianidad. Desde el rol de emprendedor, el estudiante debe determinar si es factible o no la continuación de un proyecto sobre un dispositivo para tratar el problema del sangrado en biopsias. El estudiante debe entregar al comité de ética del hospital, el comité de ética de su empresa y un grupo de inversionistas, un escrito en el que exponga su decisión y la justifique por medio de uno o dos argumentos. Así mismo, se le pide que contemple el tema de los efectos adversos del dispositivo y formule un posible contraargumento.

### Ingeniería

La prueba de Ingeniería la presentan los estudiantes del curso Dinámica de sistemas. El rol del estudiante es de un asesor externo, quien tiene que proponer una solución a una problemática en las residencias subcontratadas de una universidad. Se le presenta un modelo de referencia y se le pide que elabore uno nuevo o sugiera los cambios necesarios para su funcionamiento. Se espera que el estudiante tenga la habilidad de enriquecer la perspectiva del modelo, apoyándose en las fuentes, y proponga una solución articulada a partir de sus análisis. Como respuesta, debe entregar una presentación al Consejo de la Administración en la

---

<sup>3</sup> La razón por la que se trabajó con este departamento tiene que ver con el interés de los profesores tanto en el proceso de escritura como en la medición del pensamiento crítico a través de ella.

que explique con claridad el curso de acción por el cual se decidió y exponga con argumentos sus resultados.

### Educación

La prueba de Educación se aplica a los estudiantes de la asignatura El sentido de la educación. La situación tiene lugar en un país ficticio, en el cual una persona está interesada en fundar un centro educativo. Esta persona debe enviar su propuesta al Ministerio de Educación para su respectiva aprobación. Por esta razón, le solicita a un amigo cercano que revise el texto que elaboró. La prueba consta de documentos que dan cuenta de la organización y filosofía del colegio (por ejemplo, la misión, la visión y el currículo, entre otros), los cuales se apoyan en diferentes fuentes. Se espera que el estudiante sea capaz de identificar inconsistencias en la propuesta y señalar interpretaciones erróneas de las fuentes por parte de quien diseñó dicha propuesta. Como respuesta, el estudiante debe escribir un correo electrónico con las recomendaciones y observaciones que permitan mejorar el proyecto educativo que se quiere poner en marcha.

### Derecho

Esta prueba se aplica a estudiantes que cursan la asignatura Introducción al Derecho. Plantea una situación que involucra a dos personas que se sienten afectadas por las acciones de su contraparte. Se trata de una pareja en la que se presenta una infidelidad, y, quien la sufre, toma represalias hacia el otro a través de las redes sociales. El estudiante debe asumir el rol de asistente de Ombudsperson<sup>4</sup> y, desde esa posición, preparar un documento con una propuesta de resolución de conflictos. El curso de acción que proponga debe estar dirigido tanto a su jefe, la Ombudsperson, como a la pareja involucrada. El estudiante debe identificar las perspectivas implicadas en la situación propuesta y evaluar posibilidades para una solución, la cual se espera que no se limite a una acción punitiva.

## 5. Consideraciones sobre la evaluación

A partir del marco teórico adoptado por PALSA (Shavelson, 2018; Ulrich, 1987, 2005) y los conocimientos y experiencias de cada unidad académica, se creó una rúbrica base de cuatro grandes criterios estandarizados de evaluación, con sus

---

<sup>4</sup> Para la Universidad de los Andes Ombudsperson es una figura que ayuda a solucionar los problemas de convivencia que se presentan dentro de la institución. Esto mediante la mediación, conciliación y orientación (Universidad de los Andes, s.f.).



respectivas subcategorías, y cinco niveles de desempeño. Asimismo, cada área determinó los criterios que consideraba de mayor relevancia para la evaluación del PC, tal como se muestra en la tabla 1.

Tabla 1. Criterios de rúbrica base.

ÁREA	CRITERIOS ENFATIZADOS			
	Explorar y evaluar	Analizar perspectivas	Elaborar propuestas de soluciones	Escritura y comunicación eficaz
Educación		X	X	
Ingeniería Industrial		X	X	
Centro de Español	X			X
Derecho	X		X	

Nota. Las equis indican los criterios en los cuales enfatizan los grupos, es decir, señalan las competencias que consideran más importantes para evaluar el PC en su área.

### Diseño de rúbricas

Para apoyar el trabajo de elaboración de la rúbrica general, se revisó la rúbrica elaborada por el equipo de la Universidad Johannes Gutenberg (2018), perteneciente al proyecto iPAL. Este ejercicio permitió concretar las descripciones de los criterios o dimensiones y los niveles de desempeño.

El contenido de esta rúbrica general se estableció como guía para la elaboración de rúbricas de evaluación específicas. Esas rúbricas surgieron debido a que algunas pruebas no consideran todos los subcriterios propuestos. Además, permiten incluir otros criterios en los cuales los profesores de cada área están interesados. A continuación, se describe cada una de las dimensiones contempladas en la rúbrica general.

- a) Explorar y evaluar. La primera dimensión corresponde a la valoración y evaluación de las fuentes, ejercicio clave en las pruebas de PC. Se espera que el estudiante comprenda la información que proporcionan las diferentes fuentes, las clasifique y seleccione de acuerdo con los siguientes subcriterios:
- J) Pertinencia de la información. Se considera fundamental que el estudiante seleccione la información adecuada para comprender y analizar la situación problemática. De esta manera, un conjunto de datos o evidencias no necesariamente puede resultar pertinente para elaborar un curso de acción. En este sentido, se espera que el estudiante discrimine la información para la construcción de su respuesta y no asuma automáticamente que toda esta es útil.
  - J) Valoración de certeza y credibilidad de la información. En la selección de información se debe tener en cuenta la validez y verosimilitud de los datos y los argumentos expuestos. Adicionalmente, se espera que el estudiante identifique sesgos, es decir, información que no es objetiva o imparcial y que está condicionada por determinados intereses.
  - J) Coherencia y supuestos de los argumentos. Se busca que el estudiante cuestione los supuestos y las incoherencias de los argumentos expuestos en las fuentes proporcionadas.
  - J) Consistencia de datos y argumentos en los documentos. Diferenciar datos correctos de datos erróneos permite identificar argumentos inválidos. En este orden de ideas, el estudiante debe cuestionar inconsistencias de las fuentes en el uso de los datos. En lo posible, un estudiante que descarta una fuente debe hacer explícita la inconsistencia de datos y argumentos.
  - J) Reconocimiento de la necesidad de más información. En el examen de las fuentes, resulta necesario identificar vacíos y exponer la necesidad de información extra para proponer una respuesta. Se considera idóneo que el estudiante identifique qué información adicional necesita para una evaluación más completa del problema.
- b) Análisis de perspectivas. La evaluación del análisis de perspectivas se hace a partir de tres subcriterios. Estos exigen al estudiante examinar la manera como los involucrados se aproximan a la situación problemática y las perspectivas desde las cuales la comprenden. Además, en este análisis se espera que el estudiante pueda identificar las tensiones entre las distintas perspectivas, en cuanto aspectos como intereses, propósitos y criterios de logro.



- J) Hechos relevantes. El estudiante debe identificar los hechos que se asocian a cada perspectiva. Es decir, se espera que se concentre en examinar qué hechos o situaciones dieron lugar a que se adoptara una determinada perspectiva. De esta manera, también se establecen cuáles son los hechos más relevantes para cada actor y para los discursos a partir de los cuales se busca interpretar la situación problemática.
  - J) Actores y expertos involucrados. El estudiante debe reconocer quiénes participan directamente en la situación y cuáles son sus perspectivas; además, analizar en los discursos de los expertos las perspectivas que estos adoptan y sus aportes para una comprensión más amplia de la problemática y la elaboración de un curso de acción.
  - J) Propósitos y criterios de logro. Este criterio exige examinar si las diferentes perspectivas se oponen, parcial o radicalmente, se complementan o se aproximan, con el fin de determinar puntos de tensión o posible conciliación entre ellas. Además, sugiere considerar los aspectos que cobijan u omiten esas perspectivas. Asimismo, puntualiza en la necesidad de analizar los intereses que se defienden y los fines que se persiguen con cada perspectiva.
- c) Elaborar propuestas de soluciones. Esta dimensión comprende las habilidades del estudiante para proponer un curso de acción, apoyándose en las fuentes de información, y valorar conscientemente sus consecuencias. Además, considera la coherencia de la argumentación para sustentar su respuesta. Los subcriterios establecidos para evaluar esta dimensión son:
- J) Manejar la información para idear una solución o recomendar un curso de acción basado en un análisis de datos o información. Debe ser claro que el estudiante haya hecho un análisis consciente de los datos y la información para plantear una respuesta.
  - J) Exponer su perspectiva crítica y sus supuestos. La postura adoptada para enfrentar la situación problemática y proponer una respuesta debe ser coherente. Además, en la solución, el estudiante debe evidenciar que consideró de una forma amplia dicha situación (hechos, actores, expertos, perspectivas y criterios de logro). Adicionalmente, sus supuestos deben sustentarse y derivarse lógicamente de los análisis realizados.
  - J) Sustentar su propuesta con argumentos y contraargumentos coherentes. El desarrollo de argumentos y contraargumentos para explicar y justificar la

propuesta es fundamental. También, permite entender el hilo de pensamiento que elabora el estudiante.

- J) Considerar las consecuencias de las alternativas. Este último subcriterio es de gran importancia, pues denota que existe una comprensión superior de la problemática y una visión más responsable de la situación, dado que se reflexiona en torno a las consecuencias y los efectos de una decisión o curso de acción.
- d) Escritura y comunicación eficaz. El último criterio estándar que se estableció se relaciona con la cuestión ya abordada sobre la escritura y la comunicación. Este criterio no fue adoptado por todos los grupos, sin embargo, en él se evalúa un elemento que tiene que ver con el contexto de la prueba: Reconocimiento de la audiencia. A continuación se describen los tres subcriterios que contempla esta dimensión.
  - J) Reconocimiento de la audiencia. Se espera que el estudiante logre identificar quiénes constituyen su audiencia y cuál es el contexto de la situación, pues en la medida en que se tiene claridad sobre estos dos elementos se puede dar una comunicación eficaz.
  - J) Efectividad de la comunicación de los argumentos y las conclusiones. Este criterio apela a la capacidad que tiene el evaluado de exponer una respuesta estructurada, razonada y lógica. Es decir, no basta con exponer una decisión, una recomendación o un curso de acción, sino que resulta necesario evidenciar un ejercicio de lógica argumentativa. Esta postura parte de la idea de que el PC está relacionado con la capacidad de lograr traducir lo que se piensa a una estructura o forma textual.
  - J) Ortografía y gramática. El propósito es medir la capacidad del estudiante para expresar con claridad considerando las reglas del uso de la lengua, lo que se traduce en que su audiencia pueda tener una comprensión completa del sentido del texto o producto, pues estos elementos influyen en la efectividad de la comunicación. Algunas áreas consideran este criterio de una forma más genérica, es decir, refiriéndose al uso correcto del código o lenguaje propio de la disciplina.

## 6. Validación de los instrumentos de evaluación

Luego de diseñar las pruebas, fue necesario validarlas, es decir, determinar su adecuación a los grupos de estudiantes y examinar si, en efecto, constituían



instrumentos que dieran cuenta de los aspectos del PC que se buscan evaluar. Así mismo, las unidades académicas consideraron importante recopilar información sobre los procesos de pensamiento y las habilidades que activan los estudiantes para responder las pruebas. Con el fin de cumplir estos objetivos, se realizaron las actividades de pensar en voz alta y pilotaje.

## 6.1 Pensar en voz alta

En este ejercicio, se les solicitó a los participantes revisar la prueba y verbalizar la ruta de pensamiento que construyen a medida que avanzaban en su resolución. Se les pidió que leyeran la prueba y comentaran acerca del contexto, el rol, la tarea y los anexos o fuentes. De esta manera se recogieron impresiones y reacciones sobre la comprensión de la evaluación y su grado de dificultad. También se les solicitó una respuesta y que, oralmente, explicaran los pasos que les permitían configurarla, con lo cual se observó cómo asumen el rol asignado, juzgan las fuentes (por ejemplo, qué información consideran confiable y pertinente para dar una respuesta, y cuál descartan por sesgada o inconsistente), toman decisiones, argumentan y evalúan su propuesta. Finalmente, se formularon preguntas generales con el objetivo de obtener sugerencias o recomendaciones que ayuden a comprender y resolver la prueba.

En síntesis, esta actividad dio paso a que se ajustara la extensión de las pruebas de acuerdo con el tiempo asignado para responderlas (por ejemplo, en algunos casos fue necesario reducir el número de fuentes), se les diera a las fuentes otro orden (de manera que aquellas descartables por sesgadas o inconsistentes no quedaran al final), se incluyeran nuevas fuentes (por ejemplo, que permitieran comprender mejor el problema, dar cuenta de las perspectivas desde las cuales se aborda y establecer un diálogo entre ellas para identificar incoherencias y puntos comunes o disímiles), se perfilara mejor el rol y se diferenciara con más claridad el contexto o caso de la instrucción o tarea. Por otra parte, en la tarea de cada prueba se precisó aquello que se le solicita al estudiante: proponer una solución, exponer un curso de acción o hacer una recomendación.

## 6.2 Pilotaje

Consiste en la aplicación de las pruebas a un grupo de mayor tamaño al del ejercicio de pensamiento en voz alta. Con el fin de hacerse una idea general de aquello que pueda observarse en la aplicación definitiva, se considera importante que este grupo tenga características similares al que se enfrentará a la prueba.



El pilotaje comparte algunos objetivos con la actividad de pensamiento en voz alta, como la refinación del rol, las instrucciones y los anexos o la distinción de aspectos positivos y negativos de la prueba. Sin embargo, la calificación de las pruebas, con la implementación de la rúbrica, constituye un reto mayor, pues es necesario desarrollar convenciones específicas en torno a los criterios de evaluación que permitan abordar las diferentes respuestas de los estudiantes, con lo cual se busca comprobar la eficacia, la claridad y el potencial de la rúbrica. En síntesis, los pilotajes permitieron:

- J Recoger información sobre el comportamiento de los participantes.
- J Examinar el funcionamiento de las pruebas y de sus componentes: contexto, rol, instrucciones y fuentes.
- J Recopilar recomendaciones útiles para el mejoramiento y modificación de las pruebas (de cara a la aplicación final o una segunda fase de pilotaje, si llegara a ser necesario).
- J Evaluar la suficiencia de las pruebas respecto a la medición del PC.
- J Analizar la articulación de la rúbrica: coherencia de su estructura, precisión y pertinencia de los criterios y diferenciación o gradación de los niveles.

## 7. Hallazgos

Luego de los pilotajes, las unidades académicas se reunieron para discutir sobre el funcionamiento de las pruebas y la rúbrica. En cuanto a las pruebas, se ratificaron los acuerdos sobre su estructura, extensión, tarea que se le pide al estudiante y nivel de complejidad. Por otra parte, en algunas de las pruebas fue necesario darle un carácter más objetivo, preciso y formal al rol desde el cual se les pide a los estudiantes abordar la situación problemática, de tal manera que se favorezca un análisis alejado de condicionamientos o sesgos.

En cuanto a la rúbrica, los mayores reparos surgieron frente a algunos criterios base de evaluación. Cada lote de pruebas se revisó por lo menos dos veces, lo cual evidenció la necesidad de aclarar y diferenciar lo consignado en algunos criterios y sus niveles. En términos generales, como se mostrará a continuación, la rúbrica se amplió.

- a) Explorar y evaluar. Se dividió el subcriterio relacionado con el cuestionamiento de la información en cuanto a su credibilidad, ya que se determinó que se puede objetar la credibilidad de las fuentes por falta de autoridad y por conflicto



de intereses. Además, se especificaron acciones que permiten entender cómo es posible discutir la coherencia interna de los documentos, las inconsistencias entre documentos y los supuestos de los argumentos.

- b) Análisis de perspectivas. Para este criterio se consideró separar el análisis de perspectivas con respecto al de posiciones; así mismo, se estableció un subcriterio para cada una de estas habilidades: identificar perspectivas y describirlas, reconocer los hechos no relevantes, es decir, aquellos que no se contemplan en una perspectiva, y analizar los propósitos y criterios de logro de las perspectivas examinadas.
- c) Elaborar propuesta de solución. En este criterio se situaron todos los elementos que se usan para la descripción y explicación de una perspectiva. Esto con el fin de medir cómo en la construcción de la perspectiva y la posición propia, el estudiante contempla y presenta elementos como consecuencias, propósitos, hechos relevantes, actores involucrados y expertos.
- d) Escritura y comunicación eficaz. El cambio más representativo en este criterio fue la inclusión de una subcategoría que mida la capacidad para presentar una respuesta de forma estructurada y lógica.

## Aprendizajes y conclusiones

Este ejercicio de sistematización permitió hacer un examen de las fases del proyecto de creación y validación de pruebas estandarizadas de desempeño de PC; actividad compleja en tanto que, como se evidenció en el artículo, el proyecto se desarrolló en distintas etapas y exigió una revisión constante de sus productos. Al ofrecer con cierto grado de detalle el paso a paso del proceso de diseño de los instrumentos de evaluación, se brinda información valiosa a la comunidad académica interesada en este campo de conocimiento y a quienes están emprendiendo investigaciones en esta área. A continuación, se presentan conclusiones generales del proyecto PALSA.

El proyecto, además de generar distintos aprendizajes para sus integrantes, permitió construir consensos clave para la definición, la comprensión y la medición del PC. Una razón por la cual el grupo alcanzó esos acuerdos corresponde al gran esfuerzo hecho en la primera etapa del proyecto: la conceptualización de los elementos del PC. Al enfocarse en este trabajo de forma consciente y rigurosa, y con tiempo suficiente, se consiguió una sintonía generalizada. Sin embargo, por razones como la formación, los cursos que imparten y las expectativas sobre la prueba, no siempre los integrantes coincidieron en sus apreciaciones. Adicionalmente, vale aclarar que

el proyecto todavía no ha finalizado y aún hay aspectos en los que se puede ahondar.

Como se ha dejado ver en el artículo, la interdisciplinariedad es la columna vertebral del proyecto PALSA. Al respecto, las conclusiones más importantes se relacionan con tres ideas: el valor de la construcción conjunta de la definición de PC, la rúbrica y las cuatro pruebas; la comprensión y aceptación de las diferencias; y el aporte directo o indirecto a las disciplinas.

Lo que en principio fue el mayor reto luego se convirtió en la principal fuente de insumos para la consolidación del proyecto. Llegar a una definición de PC requirió un proceso anterior de definición de puntos en común. Por eso, para las áreas participantes es absolutamente claro que no solo se está construyendo algo novedoso, sino que, además, se construye entre todos. Aquí se puede apelar al principio de reciprocidad: cada área aportó desde su experiencia disciplinar, y ese saber fue útil para los diferentes enfoques. En palabras de uno de los participantes: “Trabajar con campos profesionales diferentes ayuda a pensar las cosas más ampliamente” (A. Mejía, comunicación personal, 5 de febrero de 2019).

De otro lado, para algunas áreas resulta reveladora la huella que el proyecto está dejando en la pedagogía y los contenidos académicos de sus clases. Es el caso, por ejemplo, de los grupos de Ingeniería y Derecho, quienes articularon o integraron las pruebas a una de sus asignaturas. Para el caso de los primeros, las reflexiones generadas en el marco de PALSA están teniendo un impacto directo en el currículo de la carrera. Para el grupo de Derecho, el proyecto de PC, en especial el enfoque ulrichiano, ha ayudado a cuestionar la formación tradicional de la disciplina, que tiende a ser, de acuerdo con los participantes de este grupo, sesgada y dicotómica en algunas ocasiones.

Para los grupos, los logros y aprendizajes no solo se dieron en términos teóricos, también lo fueron en términos prácticos respecto al área de evaluación y a los estudios de PC. Aquí es pertinente traer a colación lo dicho por una de las profesoras participantes: “Fue revelador aterrizar todo lo que se discutía en las sesiones, ver que podíamos coger esas cosas y volverlas un algo, volverlas una rúbrica, volverlas una prueba, volverlas una fuente, volverlas un algo que realmente estuviera reflejando eso que estamos discutiendo allá. [...] De lo teórico al hacer” (T. Cáceres, comunicación personal, 7 de febrero de 2019). Frente a las formas predominantes de estudio y medición del PC, se considera que uno de los avances de la propuesta de PALSA es, por un lado, no replicar las carencias de estas versiones de medición y, de otro lado, formular otras maneras de evaluar elementos centrales y de interés de las instituciones y programas académicos.



El Centro de Evaluación Educativa en su mayoría de proyectos trabaja con equipos homogéneos de especialistas de determinada área. Las condiciones particulares de este proyecto implicaron circunstancias nuevas; de ellas se desprenden los siguientes aprendizajes, los cuales, de manera realista, ofrecen un panorama sobre las limitaciones y retos de desarrollar pruebas de desempeño de esta naturaleza.

- ) La importancia de las evaluaciones en los procesos formativos. Trabajar conjuntamente con profesores de educación superior para la creación de pruebas, que tienen un impacto directo en el proceso formativo de estudiantes universitarios, va más allá de la usual implementación de pruebas sumativas cuya finalidad es comparar y observar niveles alcanzados por los evaluados. Este hecho ha permitido entender, entre otras cuestiones, la importancia de las evaluaciones en los procesos formativos y la necesidad de repensar las evaluaciones. Por ejemplo, el grupo de Ingeniería expresó que la investigación ha permitido discutir y actualizar las evaluaciones de acuerdo con los objetivos de aprendizaje.
- ) La dificultad de la adaptación de pruebas. Cuando algún profesor deseaba trabajar sobre una prueba que había diseñado previamente para sus cursos, es decir, adaptarla a los requerimientos de las pruebas estandarizadas de PC, resultaba algo complicado que aceptara y se adhiriera del todo a los acuerdos y las decisiones del grupo. Esto se debe, entre otras razones, a que la prueba original evalúa contenidos y habilidades específicas del curso, y eliminar alguno de esos elementos para darle cabida a otro que, aparentemente, no guarda relación explícita con sus clases, se consideraba problemático. En los casos en los que se planteó una prueba desde cero, es decir, una prueba completamente nueva, se consiguió integrar con mayor naturalidad los intereses del profesor y del grupo.
- ) La complejidad de abarcar las distintas competencias de PC en una prueba. El ejercicio de conceptualización llevó a identificar y querer evaluar una amplia cantidad de elementos del PC. En consecuencia, diseñar una prueba que tiene que medir tantos elementos ha sido un reto particularmente difícil. En este orden de ideas, se ha propuesto realizar múltiples pruebas en las que se evalúe solo uno o dos de los grandes criterios de la rúbrica.
- ) La dificultad de homogeneizar las condiciones de la prueba en cada disciplina. En un principio se pretendía homogeneizar diferentes aspectos de las pruebas, pero la libertad con la que evalúa cada uno de los

profesores, y las razones por las que las pruebas son importantes para sus cursos, hizo que este requerimiento fuera cada vez menos primordial. Un ejemplo de esto es que para el Centro de Español resultó imposible adaptarse al tiempo de la evaluación propuesto, por lo cual optó por una prueba no presencial.

Para terminar, cabe anotar que el grupo continúa interesado en perfeccionar sus instrumentos de evaluación, de esta manera se propone el reto de desarrollar más pruebas que activen las habilidades de PC que se quieren medir. Trabajo que, al mismo tiempo, supone una revisión y actualización constante de la rúbrica general. Además, el grupo considera que resultaría provechoso aplicar los aprendizajes de esta primera etapa de diseño y evaluación de pruebas en una segunda etapa, pues sin duda estos permiten concebir y visualizar propuestas más efectivas para evidenciar los procesos mentales y cognitivos asociados al PC.

## Bibliografía

- Butler, H. (2012). Halpern critical thinking assessment predicts real-world outcomes of critical thinking. *Applied Cognitive Psychology*, 26(5), 721-729. doi: [10.1002/acp.2851](https://doi.org/10.1002/acp.2851).
- 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>.
- Centro de Evaluación Educativa (2017). Formato único de presentación de proyectos para la convocatoria de la financiación de proyectos interdisciplinarios - 2017 (Propuesta de investigación). Bogotá: Universidad de los Andes.
- Escallón, E. y Forero, A. (2016). Aprender a escribir en la universidad, Bogotá, Colombia: Universidad de los Andes.
- Ennis, R. (2018). Critical thinking across the curriculum: A vision. *Topoi: An International Review of Philosophy*, 37(1), 165-184. doi: [10.1007/s11245-016-9401-4](https://doi.org/10.1007/s11245-016-9401-4).
- Hatcher, D. (2011). Which test? whose scores? comparing standardized critical thinking tests. *New Directions for Institutional Research*, 2011(149), 29-39. doi: [10.1002/ir.378](https://doi.org/10.1002/ir.378).



- Herde, C. N.; Wüstenberg, S. & Greiff, S. (Julio de 2016). Assessment of complex problem solving: what we know and what we don't know. *Applied Measurement in Education*, 29(4), 265-277.
- Instituto Colombiano para la Evaluación de la Educación-Icfes. (2015). Lineamientos generales para la presentación del examen de Estado SABER 11°. Recuperado de: <https://es.scribd.com/document/356017814/Lineamientos-Generales-Para-La-Presentacion-Del-Examen-de-Estado-Saber-11-2017-1>.
- Mejía, A. (2009). In just what sense should i be critical? an exploration into the notion of 'assumption' and some implications for assessment. *Studies in Philosophy and Education: An International Journal*, 28(4), 351-367. doi:[10.1007/s11217-009-9135-5](https://doi.org/10.1007/s11217-009-9135-5).
- Ministerio de Educación Nacional MEN (2011). Propuesta de lineamientos para la formación por competencias en educación superior. Recuperado de: [https://www.mineducacion.gov.co/1621/articles-261332\\_archivo\\_pdf\\_lineamientos.pdf](https://www.mineducacion.gov.co/1621/articles-261332_archivo_pdf_lineamientos.pdf).
- Ministerio de Educación Nacional MEN (2010). Sistema educativo colombiano. Recuperado de: <https://www.mineducacion.gov.co/1759/w3-article-233839.html>.
- Moore, T. (2011). Critical thinking and disciplinary thinking: a continuing debate. *Higher Education Research & Development*, 30(3), 261-274. doi: [10.1080/07294360.2010.501328](https://doi.org/10.1080/07294360.2010.501328).
- 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.; Zlatkin-Troitschanskaia, O. & Marino, J. (2017). Performance Indicators of Learning in Higher-Education Institutions: Overview of the Field. En E. Hazer Korn, H. Coates & A. Cormick (Eds.), *Research Handbook on Quality, Performance and Accountability in Higher Education*. Edward Elgar (in press).
- Ulrich, W. (1987). Critical heuristics of social systems design. *European Journal of Operational Research*, 31(3), 276-283. Recuperado de: [https://www.academia.edu/23198797/Ulrich\\_W.\\_1987\\_.Critical\\_heuristics\\_of\\_social\\_systems\\_design](https://www.academia.edu/23198797/Ulrich_W._1987_.Critical_heuristics_of_social_systems_design).

- Ulrich, W. (2005). A brief introduction to critical systems heuristics (CSH). ECOSENSUS project site, Open University, Milton Keynes. Recuperado de: [http://projects.kmi.open.ac.uk/ecosensus/publications/ulrich\\_csh\\_intro.pdf](http://projects.kmi.open.ac.uk/ecosensus/publications/ulrich_csh_intro.pdf).
- Ulrich, W. & Reynolds, M. (2010). Critical systems heuristics. En Reynold, M. & Holwell, S. (Eds.), *Systems approaches to managing change: A practical guide* (pp. 243-292). Londres: Springer.
- Universidad de los Andes. Ombudsperson. Recuperado de: <https://uniandes.edu.co/es/noticias/comunidad/oficina-ombudsperson>.
- Universidad Johannes Gutenberg, proyecto iPALs. (2018). (Rúbrica).
- Zlatkin-Troitschanskaia, O., Pant, H. A. & Coates, H. (2016). Assessing student learning outcomes in higher education: Challenges and international perspectives. *Assessment & Evaluation in Higher Education*, 41(5), 655-661. doi:[10.1080/02602938.2016.1169501](https://doi.org/10.1080/02602938.2016.1169501).

#### Forma de citar este artículo

- Crump, Sepúlveda, Fajardo y Vera (2019). Sistematización de pruebas de desempeño en pensamiento crítico: una experiencia de construcción interdisciplinar. *Revista Estudios en Educación*, Vol. 2 (2), 17-47, Santiago, Chile: Universidad Miguel de Cervantes.  
En: <http://ojs.umc.cl/index.php/estudioseneducacion/index>.

Fecha de recepción: 30/04/2019.

Fecha de aceptación: 26/06/2019.

