

IDENTIFICACIÓN DE ELEMENTOS IRRELEVANTES PARA LA COMPRESIÓN DE ÍTEMS DE UNA PRUEBA DE RAZONAMIENTO

KAREN CALVO-DÍAZ / GUANER ROJAS ROJAS / NELSON PÉREZ ROJAS / ARMANDO JOSÉ RÍOS SÁNCHEZ

Resumen:

El objetivo de este artículo es describir cuáles son los posibles elementos de comprensión que afectan la resolución de ítems nuevos en una prueba de razonamiento verbal. El punto de partida fue el reporte verbal de un grupo de 13 estudiantes que obtuvieron altas y bajas puntuaciones en la Prueba de Aptitud Académica de la Universidad de Costa Rica. Se realizó un estudio cualitativo exploratorio en el que se identificaron cinco elementos irrelevantes para la comprensión: léxico, redacción, extensión, contexto muy específico y uso de conocimiento previo, que afectan la resolución de ítems de razonamiento en contexto verbal. El reconocimiento de estos aspectos derivó de los resultados de un proyecto de investigación, que buscó determinar los elementos del enunciado, la instrucción o los distractores que dificultan la comprensión y la relación con la dificultad asignada por el examinado. Asimismo, se planteó el uso del reporte verbal como generador de evidencias de validez para una prueba de admisión.

Abstract:

The objective of this article is to describe the possible elements of comprehension that affect the response to new items on a verbal reasoning test. The starting point was the verbal report of a group of thirteen students obtaining high and low scores on an academic aptitude test at Universidad de Costa Rica. An exploratory qualitative study identified five irrelevant elements for comprehension: lexicon, rhetoric, length, very specific context, and use of previous knowledge, which affect the response to items of reasoning in the verbal context. The recognition of these aspects was a result of the research project, which aimed at determining the elements of the sentence, instructions, or detractors that hindered comprehension, and defining their relationship to the difficulty assigned by the test-taker. The suggestion was made to use the verbal report to generate evidence of validity for an admissions test.

Palabras clave: elaboración de pruebas, evaluación de pruebas, admisión a la universidad, estudiantes, razonamiento.

Keywords: test preparation, test evaluation, university admissions, students, reasoning.

Karen Calvo-Díaz, Guaner Rojas Rojas, Nelson Pérez Rojas y Armando José Ríos Sánchez: investigadores de la Universidad de Costa Rica, Instituto de Investigaciones Psicológicas. Apdo.: 11501-2060, San Pedro, Costa Rica. CE: karenalejandra.calvo@ucr.ac.cr; GUANER.ROJAS@ucr.ac.cr; nelson.perezrojas@ucr.ac.cr; jose.rios@ucr.ac.cr

Introducción

Las pruebas estandarizadas conllevan una serie de procesos que, entre otros aspectos, vigilan de manera estricta el comportamiento de un ítem, a partir de distintos parámetros estadísticos, sin menoscabo del estudio cualitativo y en apego a la medición del constructo. En este escenario, la evaluación del ítem es una de las tareas más sensibles, antes de su colocación en una prueba que, como la que aquí se presenta –Prueba de Aptitud Académica de la Universidad de Costa Rica– selecciona a los estudiantes con mayores probabilidades de éxito académico.

Las etapas por las que pasa un ítem nuevo –también denominado experimental– incluyen, por un lado, su construcción según patrones previamente establecidos, como los *Standards for educational and psychological test* (AERA/ APA/NCME, 2014) que principalmente vigilan la pertinencia del reactivo en relación con el constructo intencionado y, por otro, el juzgamiento del ítem; este proceso incluye la participación de diferentes especialistas que verifican la pertinencia del ítem respecto del constructo que mide la presentación del formato y las adecuaciones de acceso para diferentes poblaciones que realizan la prueba (Haladyna y Rodríguez, 2013; Moreno, Martínez y Muñiz, 2004).

Para el caso costarricense se conocen investigaciones previas, centradas en el análisis de ítems de banco en pruebas de admisión; su objetivo ha sido evidenciar distintos aspectos como la definición de los contextos de razonamiento (Rodríguez, Fallas y Morales, 2009), los procesos y las estrategias que emplean los estudiantes para la resolución de ítems (Jiménez, Brizuela, Pérez y Rojas, 2015), la clasificación de las categorías del constructo subyacente de una prueba de admisión (Molina y Villalobos, 2010), la predicción de la dificultad en pruebas estandarizadas de comprensión de lectura (Brizuela y Montero, 2013) y la generación de autorreportes verbales en voz alta para la identificación de procesos de razonamiento (Brizuela, Jiménez, Pérez y Rojas, 2016).

En todas estas investigaciones, sin embargo, el objeto de estudio se ha centrado en los ítems de banco y se ha excluido el análisis de los nuevos o experimentales. En este trabajo consideramos que estos últimos son material de especial importancia, en tanto que, una vez concluido su estudio y analizados los resultados de la investigación, el reactivo es susceptible a modificaciones. Esta tarea de mejora al ítem se puede ejecutar con fa-

cilidad en aquellos que aún están a prueba y que, por tanto, pueden ser modificados, previo a su ingreso definitivo en un banco.

La bibliografía sobre los conceptos de reporte verbal, protocolos de pensamiento en voz alta y entrevista cognitiva en general es muy variada y aplicada a un sinnúmero de disciplinas (Emig, 1971; Hayes, 1981). Junto al estudio de la validez de los reportes verbales (Ericsson y Simon, 1984; Russo, Johnson y Stephens, 1989; Smagorinsky, 1998), existen investigaciones basadas en la evaluación del diagnóstico cognitivo (Leighton y Gierl, 2007) en las que se indica que la verbalización puede ser asumida con la misma rigurosidad que los datos.

En cuanto a la entrevista cognitiva también ha sido una técnica que, de acuerdo con algunos investigadores (Smith y Molina, 2011), permite obtener evidencias de validez de los reactivos que constituyen una prueba, en tanto que da cuenta del reporte del proceso mental que siguen los examinados cuando ofrecen la respuesta a los ítems.

Finalmente, otras investigaciones (Brizuela, Pérez y Rojas, 2018) han propuesto una nueva metodología denominada Respuestas guiadas por el experto que, en comparación con el método tradicional de los reportes verbales, ofrece diversas ventajas para desarrollar instrumentos de evaluación educativa.

En razón de lo anterior, el presente estudio apela a la importancia de incluir ítems nuevos como objeto de investigación, pues con ello se extrae valiosa información que puede optimizar no solo el conjunto de ítems empleados para tal fin, sino, mejor aún, la construcción general de los reactivos futuros.

Aunado a lo anterior, debe señalarse que la principal motivación para el desarrollo de este estudio versa sobre dos aspectos: la importancia que tiene para una prueba de admisión el diseño de un test con ítems de alta calidad técnica y la necesidad de contar con evidencias que ayuden a predecir cómo podría comportarse un ítem nuevo en la aplicación regular de una prueba de admisión.

El objetivo de este estudio es, pues, describir cuáles son los posibles elementos de comprensión que, en una prueba de razonamiento en contexto verbal, afectan la resolución de ítems nuevos. Sumado a lo anterior, se pretende indicar cómo se puede mejorar el procedimiento para el reporte de estos aspectos y destacar qué ventajas implica para una prueba estandarizada la información sobre nuevos ítems que se genera del reporte verbal de los examinados.

Método

Se realizó un estudio cualitativo de tipo exploratorio para identificar y describir los aspectos asociados con la dificultad de comprensión de un ítem experimental. Por medio de los reportes verbales de los estudiantes, los investigadores detectaron que el léxico, la redacción, la extensión, la especificidad contextual y el conocimiento previo son los indicadores que mayores desafíos representan para la resolución del ítem y que no tienen relevancia para el constructo que mide la prueba.

Instrumento

La prueba de admisión de la Universidad de Costa Rica es de altas consecuencias, selecciona a los estudiantes con mayores posibilidades de cursar la educación superior. Se aplica anualmente y permite que, con base en los resultados de los examinados, estos se ordenen y seleccionen por nota más alta, considerando los cupos con los que cuenta la Universidad por carrera. Esta nota de admisión se calcula tomando en cuenta las materias de cuarto ciclo de la educación diversificada y el promedio de admisión, en una escala de 0 a 100, lo cual luego se transforma a una escala de 200 a 800.

La prueba está compuesta de 75 ítems de razonamiento, 42 en contexto verbal y 33 en matemático. De estos, una pequeña cantidad (8 verbales y 7 matemáticos) son ítems nuevos que se prueban después de seguir un riguroso proceso de juzgamiento; el resto de los ítems se seleccionan de un banco de la prueba.

Con este panorama, el material de trabajo para este estudio versa en el uso de los instrumentos que se describen a continuación.

Grupo de ítems nuevos

Previo a la aplicación de la entrevista, los investigadores construyeron cerca de 70 ítems de razonamiento en contexto verbal, de los cuales 41 (5 constantes y 36 variables, distribuidos en los 4 grupos de examinados) fueron elegidos considerando los siguientes criterios: *a)* pertinencia con el constructo que mide la prueba, *b)* representación de las categorías tipificadas en la tabla de especificaciones de la prueba, y *c)* posibilidades de presentar algún problema por uso de léxico específico, complejidad en la redacción o extensión. Estos ítems fueron construidos siguiendo todos los criterios técnicos y de diseño establecidos en el programa de la Prueba de

Aptitud Académica (PAA). Por tratarse de material que puede ser empleado en la prueba de admisión, es confidencial y no puede hacerse pública la transcripción de ningún ítem, pero sí de algunas de sus apreciaciones generadas del reporte verbal de los examinados. Cada uno de los 42 ítems tuvo un mismo formato compuesto por un enunciado, la instrucción o pregunta y cinco opciones de respuesta única.

Guía de entrevista

En esta guía se consignaron las siguientes preguntas para los examinados: ¿Existe algún elemento del ítem que le impidió comprenderlo? ¿Hubo algún conocimiento previo que le ayudó a definir su respuesta? ¿Hay otra opción, además de la elegida, que le parezca atractiva? ¿Considera que el formato del ítem le dificultó llegar a su respuesta? ¿Piensa que el ítem puede mejorarse en algún sentido? ¿Existe alguna estructura gramatical o palabra que no conozca?

Tabla de codificación

En este instrumento se registraron valores numéricos determinados para cada respuesta del examinado. Esta información consideró las siguientes variables: número de ítem, tipo de ítem (variable, constante), estudiante según puntuación obtenida en la prueba (alta, baja), sexo (femenino, masculino), grupo de ítems que responde (a, b, c, d), formato del ítem (abierto, cerrado), dificultad (alta, media, baja), acierto de clave (sí, no), problema identificado (léxico, redacción, extensión, especificidad contextual y conocimiento); distractores poco plausibles o problemáticos (a, b, c, d, e) y nivel de seguridad en la respuesta del estudiante (alta, baja, media).

Participantes

Se analizó el reporte verbal de un grupo de estudiantes de primer ingreso para 2017, empadronados en alguna carrera de la Universidad de Costa Rica y que consolidaron matrícula. Este grupo se eligió por pertenecer a un grupo de cercanía generacional con la población meta a la que se aplica la prueba de admisión.

Con base en estas consideraciones, se realizaron y se grabaron las entrevistas de 13 estudiantes, mayores de edad, provenientes de colegios públicos, quienes voluntariamente contribuyeron con el proyecto y firmaron un consentimiento informado de la Universidad de Costa Rica.

De los 13 estudiantes, siete (4 mujeres y 3 hombres) obtuvieron puntuaciones altas; y los otros seis (3 mujeres y 3 hombres), bajas en la Prueba de Aptitud Académica de 2016. Los 13 examinados fueron divididos en cuatro grupos, de acuerdo con la puntuación obtenida y se les presentaron de 10 a 11 ítems, para un total de 42 reportados.

A todos los participantes se les contactó vía telefónica, se les explicó el objetivo del proyecto y la duración de la entrevista. Además, se coordinó la cita presencial en la cual se leyó y se firmó el consentimiento informado.

Procedimiento y análisis

Se recolectaron y codificaron los datos de 13 entrevistas, grabadas en su totalidad, cuya duración promedio fue de 50 minutos cada una. Se procedió con la transcripción literal e integral considerando el número de ítem y la información que ofreció el examinado de manera oral, cuya tipificación derivó en la clasificación que se observa en la tabla 1.

Posteriormente, los resultados se codificaron en una hoja de cálculo, asignando los siguientes valores para cada tipo de aspecto reportado por el examinado: 0, no detecta aspecto; 1, léxico; 2, redacción; 3, extensión; 4, contexto muy específico y 5, conocimiento. Estos elementos, tipificados previamente por cuatro especialistas, coincidieron en términos generales con los reportados por los examinados.

Asimismo, se registró la distinción por tipo de ítem, de acuerdo con el formato que se le presentó a cada examinado: por un lado, el formato de opciones, o tradicional, compuesto por el encabezado, la instrucción y las opciones. De este formato se esperaba que el examinado eligiera una opción que idealmente coincidiera con la clave. Este tipo de ítem se codificó con un valor de 1.

Por otro lado, los ítems sin opciones o de respuesta abierta (Brizuela, Pérez, y Rojas, 2018) solo ofrecían el encabezado y la instrucción y se esperaba del examinado una respuesta que pudiera brindar más información sobre la mejora del ítem; una vez realizado este ejercicio, se le mostraron al entrevistado las cinco opciones para que escogiera una, la cual debería ser muy cercana a la propuesta como clave. Este ítem se codificó con un valor de 2.

Por último, también se registró la dificultad promedio que asignó cada examinado al ítem para determinar si la identificación de un aspecto de

comprensión del ítem tenía o no relación con la dificultad para resolver el ítem. Esta variable se codificó con un valor de 0 cuando no dio respuesta, 1 para dificultad baja, 2 para media, 3 para alta, y 4 cuando el examinado otorgó al ítem varias dificultades o no estaba seguro de su respuesta.

TABLA 1

Aspecto	Definición del aspecto
Léxico	Un ítem está ligado al aspecto de comprensión por léxico si por desconocimiento del significado de una o varias palabras del enunciado principal, la instrucción o las opciones, el examinado detecta una palabra que le dificulta la resolución del ítem. En algunos casos, los examinados consideran que un término es relevante para continuar con el proceso de razonamiento, o bien, cuando, aunque se produzca la respuesta del ítem, se tiene presente que desconocer el significado de una palabra es una distracción durante el proceso de resolución: por lo tanto, ello requiere generar una suposición u asociación extra para "darle sentido" al concepto
Redacción	Un ítem está ligado al aspecto de comprensión por redacción si no cumple, a criterio de los examinados, con el uso del registro del español culto escrito y normado por la Real Academia Española de la Lengua, o bien, si el estilo de redacción del enunciado principal, de la instrucción o de las opciones de respuesta, es incomprensible e interfiere con la resolución del ítem
Extensión	Un ítem está ligado al aspecto de comprensión por extensión cuando debido a la cantidad de ideas que aparecen en el enunciado principal (encabezado del ítem), se dificulta reconocer o segmentar la información. En consecuencia, esta abundancia informativa puede, en algunos casos, interferir con la resolución del ítem y se presenta cuando las opciones son demasiado extensas, a criterio del examinado, y se hace difícil discriminar entre una y otra
Especificidad	Un ítem está ligado al aspecto de comprensión por especificidad si el contexto situacional del enunciado principal es muy restringido, específico o técnico. Debido a esta razón, el contexto del ítem resulta desconocido para el examinado y dificulta o imposibilita la resolución
Conocimiento	Un ítem está ligado al aspecto de comprensión por conocimiento cuando el examinado requiere de saberes previos para responder el ítem o si se considera que tener un conocimiento cultural o intelectual permite llegar a la respuesta

Fuente: Elaboración propia.

Debe aclararse que para determinar la dificultad final del ítem, con la cual se analiza la relación entre variables como puntaje de admisión, género, problema detectado y formato de ítem, se tomó como parámetro la dificultad que la mayoría de los examinados le asignó a cada ítem. Por ejemplo, si uno lo reportaron cuatro estudiantes y tres de ellos indicaron que tenía una dificultad media, el ítem se designa con esta última; si algún otro fue reportado por seis estudiantes y tres indicaron dificultad media o alta, el ítem se declara sin consenso; o bien, si un uno fue reportado por tres estudiantes y todos le asignaron una dificultad distinta, el ítem también se declara sin consenso.

Una vez concluida esta etapa se procedió con una tercera, en la que se analizó la información a partir de la cual se estudiaron las siguientes relaciones entre variables:

- 1) La habilidad del examinado para reportar información de acuerdo con el género (mujeres-hombre) y la puntuación obtenida en la prueba de admisión (puntuaciones altas-puntuaciones bajas).
- 2) La relación entre el problema detectado y la asignación de la dificultad del ítem.
- 3) La relación del formato del ítem y el tipo de reporte de problema.

Resultados

De las 13 entrevistas realizadas se confirmaron al menos cinco elementos que incluyen distintas especificaciones. Para cada uno se estableció una definición y se enlistaron los resultados cualitativos. Debe aclararse que algunos elementos fueron reportados mínimamente por los examinados, mientras que otros fueron identificados de forma general en los reportes verbales.

Léxico

De los 42 ítems sometidos a examen, en 12 se reportaron elementos de comprensión por léxico: seis fueron señalados por estudiantes de altas puntuaciones, quienes identificaron este aspecto en 10 ítems; mientras que los otros seis los indicaron alumnos de bajas puntuaciones y reconocieron el mismo aspecto en nueve ítems. De ello se colige que no hay diferencias sustanciales entre el reporte generado por los examinados de bajas o altas puntuaciones, pues ambos grupos reconocen un número similar de

ítems con problemas léxicos. Es decir, de los 13 entrevistados solo uno no detectó un aspecto léxico en ninguno de los ítems que resolvió, mientras que el resto, al menos en uno sí lo hizo.

Los términos que resultan problematizados mayoritariamente son los sustantivos (8 ítems) seguidos de los verbos (3 ítems) y los adjetivos (2 ítems). Asimismo, se logró identificar que en el encabezado (8 ítems) fue en donde los examinados hallaron más problemas, pues justamente la mayor cantidad de texto e información se concentra en la primera parte del ítem; luego se encontraron problemas en las opciones (4 ítems) y finalmente en la instrucción (1 ítem). El campo semántico de las palabras sobre las que no se reconoció su significado fue variado y no puede indicarse que exista uno en particular que se desconozca.

Un resultado que también se extrajo de los reportes es que los 13 ítems fueron calificados con las tres dificultades: así, cuatro se designaron de dificultad baja, cuatro de media, cuatro de alta y uno donde no hubo consenso entre quienes contestaron el ítem, pues todos los examinados le asignaron una dificultad distinta. Esto podría indicar entonces que no necesariamente la presencia, en alguna parte del ítem, de un término desconocido determinaría la imposibilidad de resolverlo, pues en todos los casos los ítems obtuvieron una respuesta por parte del examinado, aunque no siempre se acertó con la clave.

Sobre esto último, cabe resaltar que del nivel de acierto de los 27 reportes generados sobre los 13 ítems identificados con este tipo de aspecto, hubo acierto en 13 casos (8 de estudiantes con altas puntuaciones y 5 con bajas), mientras que en 14 (8 de estudiantes con bajas puntuaciones y 6 de altas) hubo desacierto con respecto a la clave.

Además, se pudo distinguir que para los estudiantes con altas puntuaciones la existencia de estas palabras poco conocidas o desconocidas no representó una dificultad real para resolver el ítem, aunque sí fue un distractor, en tanto que algunos indicaron tener que “suponer” su significado para darle sentido al enunciado, a la instrucción o a los distractores. Algunos ejemplos de los reportes verbales obtenidos de los entrevistados sobre aspectos léxicos son los siguientes:

CASO 1: No. La verdad es que el lenguaje está bastante sencillo en realidad. Solo que sí hay conceptos como [se omite la palabra por razones de confidencialidad] que sí no tengo claro. Entiendo las palabras por separado, pero más o

menos los conceptos sí están un poco enredados, pero sí están bien realmente. Debería conocerlos. Pienso yo (entrevista 7, mujer, alta puntuación).

CASO 2: No la desconozco [refiriéndose a una palabra], pero sí es un toque más rebuscada [...] pero no la desconozco (entrevista 12, hombre, alta puntuación).

CASO 3: Yo marcaría la B, porque supongo que quieren saber si sé qué es [se omite la palabra por razones de confidencialidad], además de que no me dicen qué es el problema, yo lo intuía, pero no, sí, marcaría la B (entrevista 1, hombre, alta puntuación).

No obstante, hubo casos, como los siguientes, en los que se indicó la dificultad de identificar la clave por la presencia de un aspecto léxico que dentro del constructo de la prueba se considera irrelevante. Cuando se les preguntó directamente a los examinados: ¿Hay alguna palabra desconocida para usted? o ¿Cambiaría alguna palabra en el ítem?, esto fue lo que respondieron algunos:

CASO 4: De aquí la palabra [se omite la palabra por razones de confidencialidad] todavía no tengo muy claro qué es. De hecho, me ha salido en varios textos de Humanidades y no lo comprendo entonces esta palabra, no sé, tal vez sí sea necesaria para el texto, pero yo no escogería esta opción simplemente porque no entiendo qué significa esa palabra (entrevista 9, mujer, alta puntuación).

CASO 5: La C definitivamente no la entiendo porque no sé qué es [se omite la palabra por razones de confidencialidad], entonces pierde todo el sentido para mí [...].

INVESTIGADOR: ¿Considera que [se omite la palabra por razones de confidencialidad] se puede cambiar por otra palabra o que no saberla afecta la resolución del ítem?

EXAMINADO: En este caso tal vez sí no saberla afecta la resolución del ítem, sin embargo, hay palabras que uno no las sabe, pero puede tratar de deducir. [Se omite la palabra por razones de confidencialidad] me suena como a un concepto relacionado con [se omite la palabra por razones de confidencialidad]. No sé exactamente que sea, podría cambiarse por un sinónimo más concreto, más sencillo de deducir. Sin embargo, también pienso [...] que es bueno poner este tipo de palabras, así complicadas para que la gente tenga como un mayor

conocimiento, capacidad de entendimiento de todas esas cosas, o sea sí, si colocás palabras muy sencillas, cosas muy sencillas, entonces no viene siendo un examen de admisión, pues necesita tener cierto grado de dificultad, pienso yo. O sea, estaría de acuerdo con eso, creo que dejaría [se omite la palabra por razones de confidencialidad], pero por no saberla, fallo yo (entrevista 11, hombre, baja puntuación).

CASO 6: Sinceramente sí es muy técnico [el vocabulario] porque estoy viendo que hay que saber muchos conceptos, pero estos conceptos tienen más que ver, seguro, con un [se omite el contenido por razones de confidencialidad] o algo así. Tiene que tener esto en mente. Por ejemplo, usted sabe que alguien que dice una palabra compleja si no le cuesta decirlo es porque está metido en ese tema (entrevista 6, hombre, baja puntuación).

CASO 7: La palabra que no sé es [se omite la palabra por razones de confidencialidad]. Esa no sé qué es. O sea, se puede relacionar con el texto, pero en realidad no la conozco. Hasta ahora que uno la lee, verdad, nunca la había leído. Sí la puedo relacionar, pero, o sea, en mi pensar, sin saber qué es. Hay otra que dice [se omite la palabra por razones de confidencialidad]. Depende de cómo se interprete, verdad, no sabría cómo interpretarlo acá, porque en las respuestas dicen [se omite el contenido por razones de confidencialidad]. O sea, si fuese en mi caso, no sé si la pondría o la quitaría porque no sé lo que es [se omite la palabra por razones de confidencialidad] (entrevista 5, hombre, baja puntuación).

Redacción

Dos entrevistados reportaron dos ítems con dificultades respecto de la redacción, no obstante, para el resto del grupo no hubo problema con el mismo. Tampoco se indicó algún cambio específico para mejorar la redacción donde se identificó este aspecto de comprensión.

Una posible razón de por qué este problema no se registró con la misma frecuencia que el anterior es, primero, porque el ítem, previo a su exposición, conlleva una serie de revisiones por parte de expertos en lengua española que evitan, en buena medida, problemas asociados con la redacción. Por lo tanto, en esta etapa se espera hallar la menor cantidad de faltas.

Segundo, porque aunque haya aspectos de redacción susceptibles a ser mejorados, en alguna medida, el examinado asume que el material que

se le pide revisar ya está suficientemente depurado y que posiblemente si hay algún aspecto del ítem que no logra comprender o le parece confuso, podría obedecer a que su razonamiento está equivocado o no cuenta con suficiente habilidad para responderlo, tal y como se evidencia en las siguientes transcripciones:

CASO 8: [...] ahí es donde ya me enredo un poco por la pregunta. Solo me está dando el concepto, no me está dando información, pero no veo qué es lo que en realidad quiere que saque. Eso es cuando son puros punto y seguido (entrevista 6, hombre, baja puntuación).

CASO 9: Siento que el ítem lo que le falta es ser un poco más específico acá en donde dice eso de [se omite la palabra por razones de confidencialidad]. No sé, siento que está un toque como extraña la redacción.

INVESTIGADOR: ¿Cómo lo modificaría?

EXAMINADO: Es que eso es lo que estoy pensando yo pondría como [se omite el contenido por razones de confidencialidad] porque dice [se omite la palabra por razones de confidencialidad] me suena un poco extraño. Sería como [se omite la palabra por razones de confidencialidad] algo así. Y en las otras yo podría [se omite el contenido la palabra por razones de confidencialidad] algo así, es que sí está como muy al aire, pero no sé si ese es el sentido del ítem (entrevista 9, mujer, alta puntuación).

Extensión

De los 42 ítems reportados, cinco fueron señalados como problemáticos por el uso de contextos muy extensos; el reporte lo hicieron tres participantes con altas puntuaciones y dos con bajas. De estos cinco ítems, solo en uno el reporte fue hecho por dos personas. La extensión, en promedio, rondó las 8 o 10 líneas y debido a la poca cantidad de reportes, no podría indicarse que sea un problema constante, pues más bien pareciera estar relacionado con un asunto de percepción, pues al menos cuatro ítems solo fueron reportados como muy extensos por una sola persona.

Los siguientes reportes muestran que para estos examinados la extensión, más que un problema de comprensión fue un aspecto que dificultó el ítem:

CASO 10: [...] a veces, entre más largo sea, más detalles hay que considerar, tal vez eso lo hace más difícil (entrevista 2, hombre, alta puntuación).

CASO 11: Yo diría que es apropiado que sea un poco extenso [el ítem] porque si no, no tiene una suficiente información. Hay gente que le encanta lo resumido y está bien, pero, por ejemplo, a mí llegan y me dicen: Entreguen un resumen. Yo no entrego un resumen a lo que decían una página, no, no me da. Tengo que, por ejemplo, un libro y resumirlo en una página, no, no puedo [...]. Yo digo que no lo quite [refiriéndose a la extensión] porque a la vez es una prueba para la gente. Tiene que darse cuenta de que hay que estar, que es algo aparte (entrevista 2, hombre, alta puntuación).

CASO 12: Alta [refiriéndose a la dificultad que le da al ítem]. Sí tuve que leerla varias veces, es que sí cuando yo hice el examen el año pasado siempre que había una pregunta larga, así con un texto largo, siempre la dejaba de última [...]. Ajá, porque como es tan larga uno se cansa más pensando y tratando de retener todo [...]. Al final se termina confundiendo y va a tener que volverla a leer (entrevista 8, mujer, baja puntuación).

CASO 13: Sí esa sí la veo bastante larga y pierdo el hilo del tema, entonces tengo que devolverme. Entonces esa sí creo que ha sido la única que he visto bastante larga a mi parecer (entrevista 9, mujer, alta puntuación).

CASO 14: Hay unos que tienen muchas distracciones. Pero hay otros que dan más información. Entonces depende [refiriéndose a la relación entre dificultad y extensión] (entrevista 13, mujer, alta puntuación).

Especificidad

Del total de ítems reportados, cinco fueron señalados por el uso de contextos muy específicos, tres de fueron reportados por la misma persona, una mujer con altas puntuaciones; y los otros dos, por dos hombres de bajas puntuaciones. De los cinco ítems señalados con este problema, dos eran de un contexto de ciencias sociales y tres de ciencias naturales.

CASO 15: Sí explican bien [las palabras] solamente que están en un lenguaje un poco técnico [...] lo que uno espera aquí [se omite el contenido por razones de confidencialidad] entonces uno espera aquí ver algo que termine de ser lo contrario [se omite el contenido por razones de confidencialidad] solamente que aquí en estas respuestas es donde uno tiene que saber manejar el término y si no lo hace tiene que quedarse pensando bastante y los cuatro minutos

que le quedan por respuesta no dan un nivel, una dificultad alta (entrevista 6, hombre, baja puntuación).

Conocimiento

En 13 ítems se identificó un reporte positivo en cuanto al uso de conocimiento para responderlos. Quienes lo reconocieron fueron cinco estudiantes con altas puntuaciones, tres de ellos (2 mujeres y 1 hombre) indicaron la presencia de conocimiento en más de un ítem; en el caso de los examinados de bajas puntuaciones, cuatro señalaron este aspecto en al menos un ítem y dos (mujeres) lo reconocieron en más de una ocasión para ítems diferentes.

Del conjunto de los 42 ítems, los examinados contestaron, para 20 de ellos, que no se necesitaba conocimiento. Estas respuestas las ofrecieron siete estudiantes de alta puntuación (4 mujeres y 3 hombres) y cinco de baja (2 mujeres y 3 hombres), lo que demuestra nuevamente que la identificación de conocimiento no estuvo condicionada por la habilidad de los entrevistados ni por el género.

Asimismo, con el reporte verbal generado de la entrevista, se observó que muchos examinados de bajas y altas puntuaciones tenían plena conciencia de que mantenerse dentro de lo planteado por el encabezado y por la instrucción era lo ideal para contestar el ítem, pues utilizar conocimiento previo, fuera de lo que solicita el ítem, tiende a inducir a un fallo en la respuesta, tal y como se evidencia en los siguientes casos:

CASO 16: Yo siempre prefiero no basarme en la información que conozco antes, excepto si es el significado de una palabra o algo así o tal vez si son cosas básicas, trato siempre de basarme en lo que dice el texto (entrevista 2, hombre, alta puntuación).

CASO 17: [...] Uno sí sabe, pues que [se omite el contenido por razones de confidencialidad], verdad, algunas cosas, eso lo trae uno de los Estudios sociales del colegio, pero pienso que con la información que le dan aquí es suficiente para resolverlo (entrevista 12, hombre, alta puntuación).

CASO 18: Pues..., algo mínimo..., pues he estado en ferias científicas y he escuchado sobre muchos proyectos así de que [se omite el contenido por razones

de confidencialidad] o algo así. Pero yo pienso que sí es algo bastante normal [se omite la palabra por razones de confidencialidad] (entrevista 1, hombre, alta puntuación).

CASO 19: Medio [refiriéndose a la dificultad] porque sí hay que tener el conocimiento previo de que [se omite el contenido por razones de confidencialidad] ahí tal vez podría ser importante eso (entrevista 9, mujer, alta puntuación).

Debe añadirse que las dificultades que los estudiantes les asignaron a los ítems identificados con elementos de comprensión por léxico, redacción, extensión, contexto muy específico o conocimiento tienden a ser de medias a altas. Por su parte, cuando el estudiante no reporta algún aspecto en el ítem, este no suele asociarse con dificultades altas. Esto se explica por el hecho de que cuando el aspecto identificado por el examinado complica la resolución, es esperable que la dificultad que le asigne al ítem sea de media a difícil.

Discusión

Este estudio evidenció que al identificar elementos que pueden afectar la comprensión del ítem, la técnica del reporte verbal resultó ser muy útil, pues permitió recabar información acerca de aspectos concretos en los que un ítem puede ser mejorado. Como lo muestran otras investigaciones (Brizuela *et al.*, 2016) realizadas con ítems de razonamiento en contextos verbales, este tipo de reporte aporta información muy específica, pues directamente se le pregunta al examinado qué piensa sobre cierto aspecto del ítem.

Este estudio también permitió integrar el criterio de examinados que obtuvieron bajas puntuaciones en el examen de admisión, pues se incluyó este reporte en aras de contar con un panorama mucho más amplio de la población meta a la que se aplica la prueba de admisión. Quedó en evidencia que los examinados tanto de altas como de bajas puntuaciones aportan información relevante para mejorar el ítem, pero especialmente para detectar problemas léxicos.

La selección de la población, separada entre examinados que obtuvieron altas y bajas puntuaciones en la PAA en 2016, coincidió con el comportamiento esperado. Esto permite tener una idea muy general sobre cuál es la población que se espera conteste correctamente estos ítems experimentales.

Asimismo, se recabaron algunos insumos para mejorar los ítems en términos de léxico, redacción, extensión y contexto, pues los examinados realizaron recomendaciones puntuales que, ya analizadas en conjunto, con los reportes del resto de examinados y el criterio de los expertos del área verbal, representan una mejoría necesaria para comprender el ítem.

Sin embargo, se percibió que aún existe temor de parte del estudiante para reportar y “corregir” algún aspecto en el ítem. Pareciera que se da por sentado que el ítem está bien y que las posibles modificaciones que él o ella como examinados propongan no tienen suficiente peso. Aunado a ello, para el caso de los estudiantes de alta habilidad, se pudo reconocer que ellos prefieran cuestionar su propia habilidad, antes que considerar que el ítem tiene un problema. La entrevista con estudiantes ha permitido replantear también algunos otros aspectos sobre el formato de los ítems como, por ejemplo, el uso de la expresión “con certeza” y el empleo de marcaciones gráficas y resaltadores como la negrita o el subrayado. Esto motivará futuras investigaciones para analizar su pertinencia en la construcción de nuevos ítems.

Respecto de cómo se puede mejorar la metodología para obtener el reporte verbal de elementos ligados a la comprensión en los ítems, se proponen los siguientes aspectos:

- 1) Contar con tiempo suficiente para el desarrollo de la entrevista. 50 minutos podrían ser insuficientes para el reporte de un promedio de 10 ítems, si se piensa que además de reportar el problema se debe contemplar la resolución del ítem y seguir algunos protocolos como la lectura del compromiso de confidencialidad.
- 2) Repetir, para todos los examinados, las mismas preguntas de forma constante, según la guía de preguntas previamente elaborada. Ello permite que la codificación sea más completa y estimula la participación oral en los examinados que por razones de personalidad, nerviosismo, ansiedad, limitaciones de tiempo u otro aspecto no verbalizan lo suficiente.
- 3) Colocar el ítem siempre en dos formatos para su reporte: el de respuesta cerrada (tradicional o con opciones) y abierta (sin opciones), pues del primer caso se logrará que el examinado valore el ítem de forma completa y corrija las opciones que se le presentan; del segundo

caso se obtendrá más información sobre el ítem que puede, incluso, permitir replantear, mejorar o potenciar un distractor poco plausible del ítem original.

- 4) Aumentar la cantidad de entrevistados para que con la muestra se diseñe un modelo estadístico que permita contrastar los resultados de este estudio con otros insumos de índole cuantitativa.

Finalmente, cabe indicar que dentro de las ventajas que representa el uso de los reportes verbales para los ítems experimentales están conocer cuáles son las estrategias de resolución de los ítems e identificar los aspectos asociados principalmente a léxico que no suelen ser anotados en otras etapas de juzgamiento del ítem, pues no se sabe con certeza cuál es el vocabulario real que maneja un estudiante de 17 a 19 años (población meta de la prueba de admisión).

Asimismo, el reporte verbal de un ítem experimental siempre será útil tanto para la identificación de un problema asociado a la comprensión del texto como para la verificación del cumplimiento de otros aspectos que requieren el ítem y el acceso, aunque mínimo, de un sector similar a la población meta a la que se le aplica la prueba.

Referencias

- AERA/APA/NCME (2014). *Standards for Educational and Psychological Testing*, Washington, DC: American Educational Research Association/American Psychological Association/National Council on Measurement in Education (AERA/APA/NCME).
- Brizuela, Armel y Montero, Eiliana (2013). "Predicción del nivel de dificultad en una prueba estandarizada de comprensión de lectura: aportes desde la psicometría y la psicología cognitiva", *Revista Electrónica de Investigación y Evaluación Educativa*, vol. 19, núm. 2, pp. 1-23.
- Brizuela, Armel; Jiménez, Karol; Pérez, Nelson y Rojas, Guaner (2016). "Autorreportes verbales en voz alta para la identificación de procesos de razonamiento en pruebas estandarizadas", *Revista Costarricense de Psicología*, vol. 35, núm. 1, pp. 17-30.
- Brizuela, Armel; Pérez, Nelson y Rojas, Guaner (2018). "Respuestas guiadas por el experto: validación de las inferencias basadas en los procesos de respuesta", *Actualidades Investigativas en Educación*, vol. 18, núm. 3, pp. 1-21.
- Emig, Janet (1971). *The composing processes of twelfth graders*, Urbana: National Council of Teachers of English.
- Ericsson, Andres (2006). "Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks", en A. Ericsson; N. Charness; P. Feltovich y R. Hoffman (coords.), *The Cambridge Handbook of*

- Expertise and Expert Performance*, Nueva York: Cambridge University Press, pp. 223-242.
- Ericsson, Andres y Simon, Herbert (1984). *Protocol analysis. Verbal reports as data*, Cambridge: MIT Press.
- Flower, Linda y Hayes, John R. (1981). "A cognitive process theory of writing", *College Composition and Communication*, vol. 32, núm. 4, pp. 365-387.
- Haladyna, Tomas y Rodriguez, Michael (2013). *Developing and validating test items*, Nueva York: Taylor and Francis Group.
- Jiménez, Karol; Brizuela, Armel; Pérez, Nelson y Rojas, Guaner (2015). *Búsqueda de evidencias para la fundamentación teórica de la Prueba de Aptitud Académica de la Universidad de Costa Rica*, informe final, proyecto B3351, San José: Universidad de Costa Rica.
- Leighton, Jacqueline y Gierl, Mark (2007). "Verbal reports as data for cognitive diagnostic assessment", en J. Leighton y M. Gierl (coords.), *Cognitive diagnostic assessment for education. Theory and applications*, Nueva York: Cambridge University Press.
- Molina, Mauricio y Villalobos, Janneth (2010). *Análisis de la estructura subyacente de la Prueba de Aptitud Académica (PAA) de la Universidad de Costa Rica*, informe final, proyecto 723-A8-076, San José: Universidad de Costa Rica.
- Moreno, Rafael; Martínez, Rafael y Muñoz, José (2004). "Directrices para la construcción de ítems de elección múltiple", *Psicothema*, vol. 16, núm. 3, pp. 490-497.
- Rodríguez, Odir; Fallas, Selene y Morales, Saray (2009). *Destrezas cognoscitivas en los estudiantes de décimo año de educación diversificada en la resolución de ítems de analogías*, informe final, proyecto B723-A2-165, San José: Universidad de Costa Rica.
- Russo, Edward; Johnson, Eric y Stephens, Debra (1989). "The validity of verbal protocols", *Memory and Cognition*, vol. 17, núm. 6, pp. 759-769.
- Smagorinsky, Peter (1998). "Thinking and speech and protocol analysis", *Mind, Culture, and Activity*, vol. 5, núm. 3, pp. 157-177.
- Smith, Vanessa y Molina, Mauricio (2011). *La entrevista cognitiva: guía para su aplicación en la evaluación y mejoramiento de instrumentos de papel y lápiz*, serie Cuadernos metodológicos, San José: Instituto de Investigaciones Psicológicas-Universidad de Costa Rica.

Artículo recibido: 26 de noviembre de 2018

Dictaminado: 22 de abril de 2019

Segunda versión: 29 de abril de 2019

Aceptado: 29 de abril de 2019