

Autorreportes verbales en voz alta para la identificación de procesos de razonamiento en pruebas estandarizadas

Verbal Self-reporting to Identify Reasoning Processes in Standardized Tests

Armél Brizuela Rodríguez, Karol Jiménez Alfaro,
Nelson Pérez Rojas y Guaner Rojas Rojas

Universidad de Costa Rica, Costa Rica

Resumen

Los estándares actuales, para la evaluación de la calidad psicométrica de las pruebas psicológicas y educativas, estipulan que una de las evidencias requeridas que justifican las inferencias derivadas de la aplicación de un test se refiere a las estrategias para contestar a los ítems que lo componen. Por lo tanto, el objetivo del presente artículo se propone presentar los resultados de una investigación, que consistió en la ejecución de entrevistas semiestructuradas a un conjunto de 15 estudiantes universitarios de primer ingreso, cuyos reportes orales fueron analizados con el objetivo de fundamentar un conjunto de estrategias para contestar los ítems verbales de la Prueba de Aptitud Académica de la Universidad de Costa Rica, que habían sido identificadas previamente. Los resultados indican que efectivamente los participantes emplearon las estrategias propuestas, lo cual constituye una evidencia de gran importancia sobre las habilidades de razonamiento que se miden con los ítems verbales de la Prueba de Aptitud Académica. Finalmente, se concluye con una discusión sobre los resultados acerca de la utilidad de los autorreportes verbales que recaban evidencias de validez para un test y sobre futuras investigaciones en esta línea.

Palabras clave: prueba de admisión, comprensión de lectura, procesos de respuesta, reporte verbal.

Abstract

Current standards for assessing the psychometric quality of psychological and educational tests stipulate that one indication required to justify the inferences derived from the application of a test are those related to answering strategies for the test items. Thus, this article presents the results of a study that involved the execution of semi-structured interviews with a group of 15 college freshmen, whose oral reports were analyzed to provide support for a set of strategies to answer previously identified verbal items from the "Academic Aptitude Test" (Prueba de Aptitud Académica) at the Universidad de Costa Rica. The results indicate that participants actually used the proposed strategies, which is important evidence about the reasoning skills measured by the Prueba de Aptitud Académica verbal items. Finally, we conclude with a discussion of the results, the usefulness of verbal self-reports to gather evidence for test validity and future research along these lines.

Keywords: Admission exam, Reading comprehension, Response processes, Verbal report.

Armél Brizuela Rodríguez, Karol Jiménez Alfaro, Nelson Pérez Rojas y Guaner Rojas Rojas, Universidad de Costa Rica, Costa Rica.

La correspondencia en relación con este artículo se dirige a Armél Brizuela Rodríguez, Programa de la Prueba de Aptitud Académica, Instituto de Investigaciones Psicológicas, Universidad de Costa Rica. Dirección electrónica armel9@gmail.com



Los estándares actuales, para la evaluación de la calidad psicométrica de las pruebas psicológicas y educativas, estipulan que una de las evidencias requeridas que justifican las inferencias derivadas de la aplicación de un test se refiere a las estrategias para contestar a los ítems que lo componen (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014; Padilla & Benítez, 2014). Este tipo de evidencias permite conocer detalladamente la forma en que los examinados resuelven los ítems de una prueba, lo cual es fundamental para establecer si esta se adecua al propósito para el que fue creada y a su marco teórico. En este sentido, esta estrategia de validación resulta necesaria para detectar posibles fuentes de varianza irrelevante al constructo que se pretende evaluar. También, posibilita averiguar si este está apropiadamente representado en los ítems (Haladyna & Downing, 2004; Messick, 1995).

Existen diferentes metodologías para recopilar información sobre las estrategias que utilizan las personas para abordar preguntas planteadas en instrumentos de papel y lápiz (Gorin, 2006). Desde técnicas de gran precisión, como el seguimiento ocular (Allen & Horsley, 2014) hasta aquellas mediante las cuales se pueden documentar algunos contenidos mentales, que se activan durante la ejecución de una tarea (Fox, Ericsson & Best, 2011; Smith & Molina, 2011). Muchos métodos de la psicología cognitiva han mostrado ser de gran utilidad para entender los procesos mentales involucrados en la resolución de pruebas estandarizadas (Embretson & Gorin, 2001); por ejemplo, los protocolos de pensamiento en voz alta (Ericsson & Simon, 1993) constituyen una herramienta crucial para validar matrices de atributos en los modelos de diagnóstico cognitivo (Rupp, Templin & Henson, 2010; Svetina, Gorin & Tatsuoka, 2011; Wang & Gierl, 2011). Así, son muy útiles para determinar si los examinados emplean las estrategias que, según los desarrolladores de un test, son imprescindibles para resolver correctamente los ítems (Cohen & Upton, 2007; Rupp, Ferne & Choi, 2006; Schroeder, 2011). Esta técnica también suele utilizarse en las fases iniciales del desarrollo de un test (Anguas, 2001; Díaz, Cumba, Bernal & Rivera, 2008; Martínez & Nuñez, 2007; Remor, 2005). Para una aplicación de esta técnica, en Costa Rica, puede consultarse el trabajo de Villarreal (2011), en el cual se realiza una descripción exhaustiva de los procesos de resolución de una prueba de razonamiento con figuras a partir de los reportes en voz alta de los estudiantes y del criterio de expertos sobre las estrategias y las reglas que utilizaron para contestar a las preguntas de dicha prueba.

La necesidad de identificar con cierto nivel de detalle las estrategias utilizadas por los examinados para responder a los ítems de un test estandarizado es relativamente reciente en comparación con la larga trayectoria que posee este tipo de pruebas (Hunt, 2011). Esta línea de investigación se ha construido a partir de la indagación (mediante entrevistas semiestructuradas y autorreportes verbales) sobre las destrezas que utilizan las personas para resolver ítems en pruebas estandarizadas. A continuación, se exponen algunos ejemplos de la aplicación de esta técnica en el contexto de pruebas estandarizadas de comprensión lectora.

Por ejemplo, Farr, Pritchard & Smitten (1990) mediante entrevistas introspectivas y retrospectivas lograron determinar que, pese a las diferencias entre las estrategias empleadas para resolver una prueba estandarizada, los examinados se interesaban fundamentalmente por leer las preguntas primero, para luego enfocarse en el pasaje correspondiente.

AUTORREPORTES VERBALES EN PRUEBAS ESTANDARIZADAS

Por su parte, mediante la técnica de entrevista, Powers y Wilson (1995) lograron descubrir que un grupo de examinados a quienes se les presentaron ítems de comprensión lectora del Scholastic Aptitude Tests sin los respectivos textos o pasajes obtuvieron notas altas en la escala de puntuación, inclusive arriba de lo esperable por mero azar, y observaron que la estrategia más utilizada era la de examinar la consistencia entre las preguntas y reconstruir el tema original de estas. A partir de lo anterior, los autores mencionan que estos ítems no muestran una relación directa con la comprensión de lectura, porque el texto principal resulta prescindible para contestarlos correctamente.

En otro estudio, Rupp, Ferne y Choi (2006) analizaron los datos extraídos de diez entrevistas cognitivas con lectores no nativos del inglés, a los cuales se les proporcionaron tres pasajes con varias preguntas de selección única provenientes del Canadian Test of English for Scholars and Trainees (Official Languages and Bilingualism Institute, 2015) y lograron identificar las siguientes estrategias para responder a las preguntas: a) leer superficialmente el pasaje, luego las preguntas y buscar palabras claves en el texto principal para escoger la opción correcta; b) leer todo el pasaje, buscar palabras claves, leer todas las opciones, buscar términos claves en las opciones y responder a las preguntas; c) leer primero las preguntas, luego el texto principal y contestar y d) leer las preguntas y buscar palabras claves en el texto principal para ir respondiendo a las preguntas, según los términos que los lectores iban encontrando.

Con base en este breve recuento, se concluye que los autorreportes verbales en voz alta son sumamente útiles para aportar evidencias sobre el constructo medido por un test y las posibles fuentes de varianza irrelevante, que pudieran afectar la medición de este (Castillo & Padilla, 2013; Cui & Roduta, 2013; Ercikan et al., 2010; Taylor & Dionne, 2000) siempre y cuando se utilice correctamente (Cromley & Azevedo, 2006; Leighton, 2004; Leighton & Gierl, 2007a) y en un contexto relevante (Leighton & Gierl, 2007b).

El objetivo de la presente investigación fue aplicar este método para recopilar evidencias de validez sobre los procesos de respuesta, que utilizan los examinados al enfrentarse a la Prueba de Aptitud Académica (PAA) de la Universidad de Costa Rica. Este estudio se basa en la perspectiva contemporánea sobre la validez en el ámbito de las pruebas psicológicas y educativas (Borsboom, Mellenbergh & Heerden, 2004; Cronbach & Meehl, 1955; Embretson, 1996; Embretson & Gorin, 2001; Gorin, 2007; Markus & Borsboom, 2013; Messick, 1995). De acuerdo con estos enfoques, la validez de las inferencias derivadas de las puntuaciones en un test está garantizada cuando a) el constructo existe como tal y b) las variaciones en este causan las variaciones en los puntajes obtenidos en el test. Por tal motivo, un reto primordial para la toma de decisiones basada en la aplicación de pruebas estandarizadas consiste en el de contar con evidencias empíricas, que permitan vincular las respuestas a los ítems a un conjunto determinado de estrategias para resolverlos. La presente investigación brinda un aporte en esta línea de investigación sobre la comprensión de las estrategias de respuesta utilizadas para contestar a los ítems verbales de una prueba de selección a la Universidad de Costa Rica (UCR) y a la Universidad Nacional de Costa Rica (UNA).

En lo que respecta a la PAA, esta prueba se aplica anualmente a quienes desean ingresar tanto a la UCR como a la UNA. Se creó con el propósito de evaluar habilidades de razonamiento general en contextos verbales y matemáticos. A partir de varios estudios (Jiménez & Morales, 2009-2010; Molina,

Gallardo & Cordero, 2011; Montero, Villalobos & Valverde, 2007; Rojas, 2013), ha sido posible recabar evidencias sobre la capacidad predictiva de este test, las cuales indican que la PAA permite predecir de manera efectiva el rendimiento académico de los estudiantes que ingresan a la UCR. Sin embargo, de acuerdo con las perspectivas psicométricas contemporáneas (Montero, 2013), el proceso de validación de las inferencias realizadas a partir de los puntajes de un test debe incorporar evidencias de diversa índole, no solo en cuanto a su validez predictiva.

Por ello, se han realizado diversas investigaciones sobre la calidad técnica de los ítems de la PAA al utilizar técnicas tales como (1) el Modelo Lineal de Rasgo Latente para identificar posibles predictores de la dificultad de los ítems (Brizuela & Montero, 2013), (2) el modelo de Rasch para evaluar la dificultad de los ítems y su confiabilidad (Cerdas, 2011), (3) el Análisis Factorial Confirmatorio para determinar la estructura interna de la Prueba (Molina & Villalobos, 2010) y (4) las entrevistas semiestructuradas para evidenciar los procesos de resolución de las analogías en la parte verbal de la Prueba (Rodríguez, Fallas & Morales, 2009).

Así, el presente estudio se enmarca en esta línea de investigación para recopilar evidencias de validez de las inferencias, que se realizan con base en las puntuaciones derivadas del uso de la PAA. Para ello, se aplica la técnica de los autorreportes verbales en voz alta para la identificación de los procesos de razonamiento empleados en la resolución de los ítems verbales de la PAA.

Método

El diseño de la presente investigación es de tipo descriptivo y transversal, en el tanto en el que se describen las estrategias de resolución de los ítems reportadas en voz alta y los datos solamente se recogieron en una ocasión.

Participantes

El muestreo utilizado en la presente investigación fue de tipo no probabilístico y a conveniencia, porque interesó a seleccionar a estudiantes con las notas más altas en la Prueba de Aptitud Académica aplicada en el año 2014. En la investigación participaron 15 estudiantes (8 hombres y 7 mujeres) de la Universidad de Costa Rica, quienes estaban cursando el primer año en esta institución. Se utilizaron dos criterios para seleccionar a estos estudiantes: 1) haber contestado correctamente, al menos, el 90 % de las preguntas de la PAA y 2) provenir de una institución pública de educación secundaria. Todos los participantes fueron contactados por vía telefónica y antes de iniciar las entrevistas leyeron y firmaron un consentimiento informado de participación aprobado por el Comité de Ética de la UCR.

El primer criterio de selección se justifica porque para reportar en voz alta las estrategias utilizadas en la resolución de los ítems se requiere una gran habilidad verbal, puesto que el reporte se realiza de manera simultánea al proceso de respuesta. Este tipo de tareas demanda una gran cantidad de recursos cognitivos, por lo cual fue necesario garantizar que los participantes podían abordar de manera exitosa tanto la resolución de los ítems de la prueba como el reporte en voz alta de las estrategias empleadas. El segundo criterio de selección se basa en que los estudiantes provenientes de colegios públicos constituyen una gran mayoría de la población que ingresa a la UCR (Programa Estado de la Nación, 2013), por lo cual interesaba que el perfil de los participantes fuera similar al de quienes mayoritariamente realizan la PAA.

Instrumentos

Banco de ítems de la PAA: un total de 569 ítems verbales del banco de ítems de la PAA. En primer lugar, cabe señalar que los ítems utilizados en la parte de razonamiento en contexto verbal se utilizan regularmente para el proceso de admisión a la UCR y a la UNA, por lo cual son de naturaleza confidencial y no pueden ser incluidos en este trabajo. Por una parte, los ítems de comprensión de lectura se componen de un párrafo (entre 1 y 13 renglones de extensión), una pregunta que debe contestarse principalmente con base en la información planteada en este y cinco opciones de respuesta (cuatro distractores y una opción correcta). Por otra parte, los ejercicios de completar oraciones consisten en párrafos (entre 1 y 5 renglones) en los que se eliminan una o varias palabras, las cuales aparecen en las opciones de respuesta junto con otras que no se adecuan al texto. En los ítems de completar oraciones, se le solicita a los examinados que elijan la opción en la que aparecen las palabras que completan correctamente el sentido del párrafo, mientras que en los de comprensión de lectura deben seleccionar una respuesta que se derive del párrafo.

Test para la entrevista: una prueba conformada por un subconjunto de 17 ítems de banco de la PAA.

Guía de entrevista: una guía de entrevista para obtener información específica sobre las estrategias de resolución empleadas, la cual contenía preguntas como las siguientes: ¿Cómo llegó a esa respuesta?, ¿qué información utilizó para contestar el ítem?, ¿por qué marcó esa opción?, ¿hay alguna palabra o frase particularmente útil para contestar el ítem?, entre otras.

Procedimientos

Los autores de la presente investigación resolvieron 569 ítems verbales del banco de ítems con que se cuenta para ensamblar los formularios de la PAA. De manera simultánea, se realizó una revisión bibliográfica sobre teorías cognitivas, lingüísticas y pedagógicas y de cómo las personas entienden los textos escritos (Belinchón, Igoa & Rivière, 1998; Gutiérrez, 1999; Kintsch, 1998; McNamara & Magliano, 2009; Murphy, 2003). Con base en el criterio experto de los investigadores (quienes han trabajado en el desarrollo de la prueba durante varios años) y en los aportes teóricos señalados, se elaboró un sistema de categorías basado en los procesos de comprensión lectora documentados en la literatura y que, a su vez, respondiera a las estrategias específicas requeridas para resolver los ítems verbales de esta prueba (Parafrasear, Identificar marcadores, Oponer y Reducir), cuyas definiciones se incluyen en la Tabla 1.

A. BRIZUELA, K. JIMÉNEZ, N. PÉREZ Y G. ROJAS

Tabla 1

Categorías de análisis

Categoría	Definición
Parafrasear	La paráfrasis se refiere a una relación entre conceptos o elementos discursivos, tales como la sinonimia o las operaciones de reformulación, respectivamente. La sinonimia constituye una relación entre dos signos con expresiones diferentes que están asociados con el mismo significado (Del Teso & Núñez, 1996). La reformulación es un procedimiento de organización mediante el cual un hablante (o el autor de un texto) retoma un elemento discursivo anterior para reinterpretarlo y presentarlo de manera distinta (Garcés, 2006). En el contexto de la PAA, los participantes deben realizar dos procesos, dependiendo del tipo de paráfrasis presente en los ítems: uno, percatarse de que una o dos palabras del texto principal son sinónimas (u paráfrasis en algún sentido) entre sí o respecto de la respuesta correcta; dos, identificar la igualdad semántica entre una o varias proposiciones del texto principal y la opción correcta.
Identificar marcadores	Su nombre se debe a la importancia que tienen en ella los recursos idiomáticos denominados marcadores discursivos. Los marcadores discursivos son unidades lingüísticas invariables cuya función consiste en guiar las inferencias que se realizan en la comunicación, de acuerdo con sus propiedades morfosintácticas, semánticas o pragmáticas (Martín & Portolés, 2000). Además, como indica Sánchez (2005, p. 172), “actúan como señales de cómo debe procesarse una porción informativa en relación con el resto del texto y el contexto, pero carecen del significado léxico que poseen otras unidades (como ‘lámpara’, ‘bonito’ o ‘caminar’)”. En la PAA, las personas evaluadas deben utilizar los marcadores discursivos presentes en el enunciado principal para determinar la relación entre dos porciones ese texto e identificar la clave del ítem; en otros casos, quienes resuelven el ítem deben encontrar la relación existente entre dos porciones del texto principal y encontrar, entre las opciones de respuesta, el marcador que mejor representa dicha relación.
Oponer	En una oposición se presentan dos unidades lingüísticas, léxicas o textuales cuyos significados son contrapuestos. Una de las particularidades de la oposición consiste en que los elementos implicados mantienen una relación de cercanía a pesar de ser contrarios: están relacionados, pero se encuentran en los extremos de un eje de significado (Duță, 2009,). Las relaciones de oposición se suelen dividir en tres tipos (Duță, 2009; Lyons, 1968): complementariedad, antonimia y reciprocidad. En el contexto de la PAA, las oposiciones pueden darse sin la necesidad de marcadores discursivos y el evaluado debe realizar dos procesos: uno, percatarse de que una o dos palabras del texto principal son antónimas (u opuestas en algún sentido) entre sí o respecto de la respuesta correcta; dos, identificar el contraste o contradicción entre dos afirmaciones.
Reducir	Esta corresponde a la identificación del tema, que teóricamente se ha definido como una palabra o un sintagma que explica sobre qué trata el texto (Aulls, 1978). La coherencia y la cohesión son propiedades que contribuyen con la unidad del texto; dicha unidad, a su vez, es necesaria para que el lector pueda determinar cuál es el tema del texto: debe existir un eje que permita la comprensión global de la información presentada. En la PAA, los evaluados deben realizar dos posibles tareas: uno, identificar la síntesis de la información expuesta en el texto principal en una proposición; dos, descubrir un elemento en común entre los significados de varias palabras, frases u oraciones.

AUTORREPORTES VERBALES EN PRUEBAS ESTANDARIZADAS

Posteriormente, los 569 ítems fueron clasificados de acuerdo con las categorías definidas. Cabe señalar que las categorías con las que se clasificaron los ítems, se derivan directamente de los ítems del banco, por lo que fue imprescindible resolver la gran mayoría de los ítems del banco verbal con el objetivo de abarcar la totalidad de las estrategias de resolución involucradas.

Posteriormente, de los 569 ítems del banco, se eligieron 17 ítems representativos de las siguientes categorías de la Tabla 1: reducir, oponer, identificar marcador y parafrasear. Estas categorías fueron elegidas por ser de particular interés para los desarrolladores de la PRUEBA, quienes consideran que estas son inherentes a la comprensión lectora. A cada participante se le administraron los 17 ítems (en un orden diferente y al azar para cada uno) para que los resolviera en voz alta, esto es, para que reportara todo lo que pensaba mientras intentaba seleccionar una respuesta (reporte verbal concurrente); los ítems se administraron de uno en uno, en folios separados. Una vez elegida la opción, a los participantes se les realizaban las preguntas señaladas anteriormente para obtener más información sobre las estrategias empleadas (reporte verbal retrospectivo). Estas entrevistas fueron aplicadas entre el 11 de septiembre y el 4 de noviembre de 2013.

Finalmente, las entrevistas fueron codificadas de manera independiente, por dos de los autores, en términos de si para cada uno de los 17 ítems se evidenciaba que los estudiantes hubieran utilizado una estrategia relacionada con la categoría a la que fue inicialmente asignado cada ítem; es decir, la codificación (0 = no se utilizó la estrategia esperada y 1 = se utilizó la estrategia esperada) se realizó con el objetivo de corroborar que la clasificación hecha por los autores estuviera respaldada por la forma en que los estudiantes contestan a los ítems verbales de la PAA. Para evaluar la confiabilidad de los jueces, se utilizó el coeficiente kappa de Cohen para cada categoría.

Resultados

El coeficiente kappa de Cohen para cada categoría se expone en la Tabla 2. La columna “Antes” presenta el kappa obtenido de una primera codificación independiente de los autores. Una vez observados estos valores, los jueces se reunieron para discutir las discrepancias en la codificación, que generaron valores bajos en el coeficiente kappa. Una vez realizada esta discusión, los autores volvieron a codificar las entrevistas de acuerdo con la misma consigna y se volvió a calcular el coeficiente kappa (columna “Después”).

Tabla 2
Concordancia entre codificadores

Categoría	Antes	Después
Parafrasear	.196	.523
Identificar marcadores	.624	.872
Oponer	.347	.566
Reducir	.773	.813
Total	.539	.723

Para la interpretación del coeficiente kappa de Cohen, se utilizaron los parámetros recomendados en la literatura (Landis & Koch, 1977). De acuerdo con Landis y Koch (1977), el coeficiente kappa se puede interpretar con los siguientes parámetros: pobre (menor a 0), ligero (0 – 0.20), justo (0.21 – 0.40), moderado (0.41 – 0.60), sustancial (0.61 – 0.80) y casi perfecto (0.81–1).

Como se puede observar en la Tabla 1, posteriormente a la reunión de los jueces, el coeficiente kappa para las categorías “parafrasear” y “oponer” es moderado, mientras que para las categorías “identificar marcador” y “reducir” resulta casi perfecto. De esta manera, las cuatro categorías expuestas se consideran confirmadas mediante el acuerdo entre jueces, al menos, en un nivel moderado. A continuación, se muestran ejemplos de las frases registradas en las entrevistas que ratifican el uso de las estrategias de resolución propuestas por los autores.

Parafrasear

Estas frases se encuentran en los reportes que dieron los participantes posteriormente a la resolución de cada ítem: “lo mismo” y “lo que dice arriba” se refieren al enunciado principal, o una porción de este, y son el medio utilizado por los participantes para justificar la escogencia de una opción determinada.

1. Es lo mismo, pero más resumido.
2. Es una traducción de lo que dice arriba.
3. La b sería lo mismo nada más que al revés.
4. Es lo que dice arriba con palabras distintas.

Identificar marcadores

En estos ejemplos, los participantes reportan el marcador discursivo específico que les permite reconocer la respuesta correcta para cada ítem: “es tan”, “como”, “si” y “aunque”. Además, el registro de los marcadores de los ejemplos 5), 6) y 7) está acompañado de una explicación de cómo ese elemento específico contribuye con la resolución del ítem: “está comparando”, “implica un símil” y “me remite a una condición”.

5. “Es tan”... está comparando.
6. Tenía dudas por el “como”, implica símil.
7. El elemento “si” me remite a una condición.
8. Hay una palabra importante; es el “aunque”.

Oponer

Como se puede apreciar en estos ejemplos, el reporte de los participantes suele implicar palabras directamente relacionadas con el nombre mismo de la categoría: “opuesto”, “contradicción”, “antagónico”, como en 9), 10) y 11); sin embargo, hay explicaciones más amplias que exponen la relación existente entre los elementos discursivos necesarios para resolver el ítem, tal como en la frase 12).

AUTORREPORTES VERBALES EN PRUEBAS ESTANDARIZADAS

9. Tiene que ser algo opuesto.
10. Nos habla de una contradicción.
11. Estaba buscando algo antagónico.
12. Está dando un factor negativo, pero también causa algo bueno.

Reducir

En el reporte de la categoría “reducir”, en algunos casos, los participantes reconocen directamente qué se les pide, como en 13); otras veces utilizan frases que, precisamente, se relacionan con la definición teórica de tema, como en 14); en otros casos, confirman que deben utilizar un proceso definido para los ítems de esta categoría, como en 15) y 16).

13. Es el tema principal.
14. De qué me habla el texto.
15. Busco algo que pueda englobar todas estas condiciones.
16. Me piden la frase que relacione todas estas afirmaciones que vienen después, que son ejemplos del mismo concepto.

Discusión

La técnica de los autorreportes en voz alta demostró ser de gran utilidad para indagar sobre los procesos de respuesta a los ítems. Tal y como ha sido de gran utilidad para investigar otras pruebas, mediante esta técnica se evidenció que las estrategias planteadas por los autores las utilizan los examinados. A la luz de estos resultados, los autores recomiendan el uso de esta técnica como una primera aproximación para conocer la forma en que las personas se plantean la tarea de contestar los ítems de un test.

En esta investigación se presentó evidencia inicial sobre las estrategias utilizadas para contestar los ítems verbales de la PAA. En este sentido, se aportaron evidencias sobre los procesos de razonamiento que utilizan los examinados para contestar a este tipo de preguntas: identificar equivalencias u oposiciones semánticas entre porciones de un enunciado o entre un enunciado y la opción de respuesta correcta; identificar la síntesis de la información expuesta en el texto principal en una proposición o descubrir un elemento en común entre los significados de varias palabras, frases u oraciones y utilizar los marcadores discursivos presentes en el enunciado principal para determinar la relación entre dos porciones ese texto e identificar la clave del ítem.

Los resultados también ofrecen una luz sobre las estrategias de razonamiento empleadas en tareas verbales, las cuales suelen denominarse tradicionalmente como “deducción”, “inducción” o “analogía”. En esta investigación se desglosaron detalladamente qué tipo de estrategias se emplean a la hora de abordar los ítems verbales de la Prueba de Aptitud Académica, lo cual puede servir como guía para quienes se dedican a la construcción y el uso de pruebas estandarizadas relacionadas con habilidades verbales desean profundizar en el constructo medido por estas. En este sentido, los actuales estándares para la medición psicológica y educativa (American Educational Research Association, American Psychological

Association & National Council on Measurement in Education, 2014) plantean la necesidad de recabar evidencias de validez, las cuales son necesarias para fundamentar las inferencias, que se realizan con base en la aplicación de una prueba.

Para futuras investigaciones, este trabajo se podría complementar con el uso de técnicas analíticas cuantitativas, tales como el Análisis Factorial Confirmatorio, para recabar evidencias sobre si los ítems efectivamente se agrupan de acuerdo con las estrategias identificadas por los autores en esta prueba.

Referencias

- AERA (American Educational Research Association), APA (American Psychological Association) y NCME (National Council on Measurement in Education). (2014). *Standards for Educational and Psychological Testing*. Washington, Estados Unidos: American Educational Research Association.
- Allen, B. & Horsley, M. (2014). A New Approach to Cognitive Metrics: Analysing the Visual Mechanics of Comprehension Using Eye-Tracking Data in Student Completion of High-Stakes Testing Evaluation. En M. Horsley, M. Eliot, B. Allen & R. Reilly (Eds.), *Current Trends in Eye Tracking Research* (pp. 287-296). Suiza: Springer.
- Anguas, A. (2001). Identificación y validación del significado subjetivo en México: Fundamentos para el desarrollo de un instrumento de medición. *Interamerican Journal of Psychology*, 35(1), 163-183.
- Aulls, M. (1978). *Developmental and remedial reading in the middle grades*. Boston: Allyn and Bacon.
- Belinchón, M., Igoa, J. & Rivière, A. (1998). *Psicología del lenguaje: Investigación y teoría*. España: Editorial Trotta.
- Borsboom, D., Mellenbergh, G.J. & Heerden, J. (2004). The Concept of Validity. *Psychological Review*, 111(4), 1061-1071.
- Brizuela, A. & Montero, E. (2013). Predicción del nivel de dificultad en una prueba estandarizada de comprensión de lectura: aportes desde la psicometría y la psicología cognitiva. *RELIEVE*, 19(2), 1-23.
- Castillo, M. & Padilla, J. (2013). How cognitive interviewing can provide validity evidence of the response processes to scale items. *Social indicators research*, 114(3), 963-975.
- Cerdas, D. (2011). *Construcción de una tabla de especificaciones para la parte matemática de la prueba de aptitud académica de la Universidad de Costa Rica, mediante la aplicación del modelo de Rasch* (Tesis de maestría). Costa Rica: Universidad de Costa Rica.
- Cohen, A. & Upton, T. (2007). 'I want to go back to the text': Response Strategies on the Reading Subtest of the New TOEFL. *Language Testing*, 24(2), 209-250.
- Cromley, J. & Azevedo, R. (2006). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning*, 1(3), 229-247.
- Cronbach, L.J. & Meehl, P.E. (1955). Construct Validity in Psychological Test. *Psychological Bulletin*, 52(2), 281-302.

AUTORREPORTES VERBALES EN PRUEBAS ESTANDARIZADAS

- Cui, Y. & Roduta, M. (2013). Validating Student Score Inferences With Person-Fit Statistic and Verbal Reports: A Person-Fit Study for Cognitive Diagnostic Assessment. *Educational Measurement: Issues and Practice*, 32(1), 34-42.
- Del Teso, E. & Núñez, R. (1996). *Semántica y pragmática del texto común*. España: Cátedra.
- Díaz, M., Cumba, E., Bernal, G., & Rivera, C. (2008). Desarrollo y propiedades psicométricas de la escala de autoeficacia para la depresión en adolescentes (EADA). *Interamerican Journal of Psychology*, 42(2), 218-227.
- Duță, O. (2009). *Semántica Léxica y oposiciones de sentido: un enfoque teórico*. Annals of the University of Craiova. Series Philology. Linguistics. Recuperado de www.cceol.com
- Embretson, S. & Gorin, J. (2001). Improving Construct Validity Cognitive Psychology Principles. *Journal of Educational Measurement*, 38(4), 343-368.
- Embretson, S. E. (1996). The New Rules of Measurement. *Psychological Assessment*, 8(4), 341-349.
- Ercikan, K., Arim, R., Law, D., Domene, J., Gagnon, F., & Lacroix, S. (2010). Application of think aloud protocols for examining and confirming sources of differential item functioning identified by expert reviews. *Educational Measurement: Issues and Practice*, 29(2), 24-35.
- Eriasson, K. & Simon, H. (1993). *Protocol Analysis*. Estados Unidos: MIT Press.
- Farr, R., Pritchard, R. & Smitten, B. (1990). A Description of What Happens When an Examinee Takes a Multiple-Choice Reading Comprehension Test. *Journal of Educational Measurement*, 27(3), 209-226.
- Fox, M., Ericsson, K. & Best, R. (2011). Do Procedures for Verbal Reporting of Thinking Have to Be Reactive? A Meta-Analysis and Recommendations for Best Reporting Methods. *Psychological Bulletin*, 137(2), 316-344.
- Garcés, M. (2006). *Las operaciones de reformulación*. Actas del XXXV Simposio Internacional de la Sociedad Española de Lingüística, España, 84-690-3383-2
- Gorin, J. (2006). Test Design with Cognition in Mind. *Educational Measurement: Issues and Practice*, 25(4), 21-35.
- Gorin, J. S. (2007). Reconsidering Issues in Validity Theory. *Educational Researcher*, 36(8), 456-462.
- Gutiérrez, M. (1999). Inferencias en la comprensión del lenguaje. En M. Vega & F. Cuetos (Eds.), *Psicolingüística del español* (pp. 231-270). España: Editorial Trotta.
- Haladyna, T. & Downing, S. (2004). Construct-Irrelevant Variance in High-Stakes Testing. *Educational Measurement: Issues and Practice*, 23(1), 17-27.
- Hunt, E. (2011). *Human Intelligence*. Estados Unidos: Cambridge University Press.
- Jiménez, K. & Morales, E. (2009-2010). Validez predictiva del Promedio de Admisión de la Universidad de Costa Rica y sus componentes. *Actualidades en psicología*, 23-24(110-111), 21-55.
- Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. Estados Unidos: Cambridge University Press.

-
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Leighton, J. & Gierl, M. (2007a). Verbal Reports as Data for Cognitive Diagnostic Assessment. En J. Leighton y M. Gierl (Eds.), *Cognitive Diagnostic Assessment for Education* (pp. 146-172). Estados Unidos: Cambridge University Press.
- Leighton, J. & Gierl, M. (2007b). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice*, 26(2), 3-16.
- Leighton, J. (2004). Avoiding misconception, misuse, and missed opportunities: The collection of verbal reports in educational achievement testing. *Educational Measurement: Issues and Practice*, 23(4), 6-15.
- Lyons, J. (1968). *Introduction to Theoretical Linguistics*. Reino Unido: University Press.
- Markus, K. & Boorsboom, D. (2013). *Frontiers of Test Validity: Measurement, Causation, and Meaning*. Estados Unidos: Routledge.
- Martin, M. & Portolés, J. (2000). Los marcadores del discurso. En I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (pp. 4051-4213). España: Espasa-Calpe.
- Martínez, C. & Nuñez, C. (2007). Entrevista de Prototipos de Apego Adulto (EPAA): Propiedades psicométricas de su versión en Chile. *Interamerican Journal of Psychology*, 41(3), 261-274.
- McNamara, D. & Magliano, J. (2009). Towards a comprehensive model of comprehension. En B. Ross (Ed.), *The psychology of learning and motivation* (pp. 297-384). Estados Unidos: Academic Press.
- Messick, S. (1995). Validity of Psychological Assessment. *American Psychologist*, 50(9), 741-749.
- Molina, M. & Villalobos, J. (2010). *Análisis de la estructura subyacente de la Prueba de Aptitud Académica (PAA) de la Universidad de Costa Rica* [Reporte de investigación 723-A8-076]. Costa Rica: Universidad de Costa Rica.
- Molina, M., Gallardo, E. & Cordero, R. (2011). *Tiempo para graduarse en la Universidad de Costa Rica: Factores Socioeconómicos y académicos asociados* [Reporte de investigación del proyecto B1075]. Costa Rica: Universidad de Costa Rica.
- Montero, E. (2013). Referentes conceptuales y metodológicos sobre la noción moderna de validez de instrumentos de medición: implicaciones para el caso de personas con necesidades educativas especiales. *Actualidades en Psicología*, 27(114), 113-128.
- Montero, E., Villalobos, J. & Valverde, A. (2007). Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico en la Universidad de Costa Rica: Un análisis multinivel. *RELIEVE*, 13(2), 215-234.
- Murphy, L. (2003). *Semantic Relations and the Lexicon*. Reino Unido: Cambridge University Press.
- Official Languages and Bilingualism Institute. (2015). *Canadian Test of English for Scholars and Trainees*. Canadá: University of Ottawa.

AUTORREPORTES VERBALES EN PRUEBAS ESTANDARIZADAS

- Padilla, J. & Benítez, I. (2014). Validity Evidence Based on Response Processes. *Psicothema*, 26(1), 136-144.
- Powers, D. & Wilson, S. (1995). Answering the New SAT Reading Comprehension Questions without the Passages. *Journal of Educational Measurement*, 32(2), 105-129.
- Programa Estado de la Nación. (2013). *Cuarto Informe Estado de la Educación. Costa Rica: Programa Estado de la Nación*. Recuperado de <http://www.estadonacion.or.cr/estado-educacion/educacion-informe-ultimo>
- Remor, E. (2005). Desarrollo de una medida específica para la evaluación de la calidad de vida en pacientes adultos viviendo con hemofilia en América Latina: el Hemolatin-QoL. *Revista Interamericana de Psicología*, 39(2), 211-220.
- Rodríguez, M., Fallas, S. & Morales, S. (2009). *Destrezas cognoscitivas en los estudiantes de décimo año de Educación Diversificada en la resolución de ítemes de analogías* [Informe final de investigación N° 723-A2-165]. Costa Rica: Universidad de Costa Rica.
- Rojas, L. (2013). Validez predictiva de los componentes del promedio de admisión a la universidad de costa rica utilizando el género y el tipo de colegio como variables control. *Actualidades Investigativas en Educación*, 13(1), 45-69.
- Rupp, A., Ferne, T. & Choi, H. (2006). How Assessing Reading Comprehension with Multiple-Choice Questions Shapes the Construct: a Cognitive Processing Perspective. *Language Testing*, 23(4), 441-474.
- Rupp, A., Templin, J. & Henson, R. (2010). *Diagnostic Measurement: Theory, Methods, and Applications*. Estados Unidos: The Guilford Press.
- Sánchez, A. (2005). Los conectores discursivos: su empleo en redacciones de estudiantes universitarios costarricenses. *Revista de Filología y Lingüística*, 31(2), 169-199.
- Schroeder, S. (2011). What Readers Have and Do: Effects of Student's Verbal Ability and Reading Time Components on Comprehension With and Without Text Availability. *Journal of Educational Psychology*, 103(4), 877-896.
- Smith, V. & Molina, M. (2011). *La entrevista cognitiva: Guía para su aplicación en la evaluación y mejoramiento de instrumentos de papel y lápiz*. San José, Costa Rica: Instituto de Investigaciones Psicológicas.
- Svetina, D., Gorin, J. & Tatsuoka, K. (2011). Defining and Comparing the Reading Comprehension Construct: A Cognitive-Psychometric Modeling Approach. *International Journal of Testing*, 11, 1-23.
- Taylor, K. & Dionne, J. (2000). Accessing problem-solving strategy knowledge: The complementary use of concurrent verbal protocols and retrospective debriefing. *Journal of Educational Psychology*, 92(3), 413.
- Villarreal, M. (2011). *Construcción de un modelo psicométrico cognitivo para una prueba de inteligencia fluida* (Tesis de maestría). Costa Rica: Universidad de Costa Rica.
- Wang, C., & Gierl, M. (2011). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in critical reading. *Journal of Educational Measurement*, 48(2), 165-187.

A. BRIZUELA, K. JIMÉNEZ, N. PÉREZ Y G. ROJAS

Recibido: 10 de febrero de 2016
Revisión recibida: 22 de abril de 2016
Aceptado: 27 de mayo de 2016

Reseña de los autores y la autora

Guaner Rojas Rojas es Doctor en Metodología de las Ciencias del Comportamiento y de la Salud con énfasis en medición y especialidad en Modelos de Diagnóstico Cognitivo por la Universidad Autónoma de Madrid. Funge como coordinador académico del Programa de la Prueba de Aptitud Académica adscrito al Instituto de Investigaciones Psicológicas de la Universidad de Costa Rica. Labora como docente e investigador en diversas unidades académicas de la Universidad de Costa Rica, tales como la Escuela de Estadística y el Doctorado en Educación, en donde imparte cursos de metodología de la investigación, medición y modelos probabilísticos. También, ha realizado investigaciones sobre el razonamiento en contextos matemáticos.

Karol Jiménez Alfaro es Máster en Matemática con énfasis en Matemática Educativa de la Universidad de Costa Rica. Se desempeña como investigadora y desarrolladora de la prueba del Programa de la Prueba de Aptitud Académica adscrito al Instituto de Investigaciones Psicológicas de la Universidad de Costa Rica. Ha laborado como docente en la Escuela de Matemática de la Universidad de Costa Rica impartiendo cursos de matemática elemental y elaborando materiales didácticos para la enseñanza de la matemática. Ha realizado investigaciones sobre la validez predictiva y el razonamiento en contextos matemáticos.

Nelson Pérez Rojas es bachiller en Filología Española de la Universidad de Costa Rica. Trabaja como investigador y desarrollador de prueba del Programa de la Prueba de Aptitud Académica adscrito al Instituto de Investigaciones Psicológicas de la Universidad de Costa Rica. Ha laborado como docente en diversas universidades impartiendo cursos de sociolingüística, psicolingüística y comunicación escrita. Ha realizado investigaciones sobre la medición de la comprensión lectora, el razonamiento en contextos verbales y el proceso de desplazamiento y la revitalización de la lengua brunca desde una perspectiva endocomunitaria.

Armel Brizuela Rodríguez es Máster en Investigación Psicológica de la Universidad de Costa Rica. Funge como investigador y desarrollador de prueba del Programa de la Prueba de Aptitud Académica adscrito al Instituto de Investigaciones Psicológicas de la Universidad de Costa Rica. Ha laborado como docente en la Universidad de Costa Rica impartiendo cursos de psicometría y comunicación escrita. Ha realizado investigaciones sobre la medición de la comprensión lectora, el razonamiento en contextos verbales y la elaboración de pruebas de expresión escrita.

Publicado en línea: 29 de junio de 2016