



Baremación de una prueba estandarizada de resúmenes (RESUMeV) para los niveles de 4º y 6º de educación primaria

Primer premio de la XXI edición del Premio de Psicología Aplicada “Rafael Burgaleta” 2014

José A. León^{a*}, José D. Moreno^a, Lorena A. M. Arnal^a, Inmaculada Escudero^b y Ricardo Olmos^a

^aUniversidad Autónoma de Madrid, España

^bUniversidad Nacional de Educación a Distancia (UNED), España

INFORMACIÓN ARTÍCULO

Manuscrito recibido: 20/07/2014

Revisión recibida: 11/09/2014

Aceptado: 08/12/2014

Palabras clave:

Evaluación de resúmenes
Criterios de fiabilidad y validez
Evaluación de la comprensión
Baremación
Alumnos de primaria

Keywords:

Reading comprehension
Validity
Reading assessment
Primary school

RESUMEN

El objetivo del estudio fue el de mostrar evidencias de fiabilidad y validez para una prueba de resúmenes (RESUMeV) que evalúa el grado de comprensión lectora en alumnos de cuarto y sexto de primaria. Participaron un total de 528 estudiantes de primaria, de los cuales 236 fueron estudiantes de 4º de primaria (9 y 10 años) y 292 a 6º de primaria (11 a 13 años). Todos estos alumnos procedían de 21 centros escolares. Para evaluar la consistencia interna, se calculó el alfa de Cronbach en los criterios de evaluación de resúmenes (contenido, coherencia y estilo) y su índice de homogeneidad (H_j). La validez se evaluó mediante la comparación entre niveles académicos. En cuanto a los resultados, se obtuvieron índices elevados significativos de fiabilidad y de validez. Los resultados proporcionan evidencia empírica para la validez de la prueba.

© 2014 Colegio Oficial de Psicólogos de Madrid. Producido por Elsevier España, S.L.U. Todos los derechos reservados.

Scaling of a standardized summary test (RESUMeV) for two primary school grades

ABSTRACT

The purpose of this work was to collect construct and criterion-related evidence of validity for a summary test (RESUMeV) designed to assess students from fourth and sixth grade. The sample of this study consisted of 528 children, 236 from fourth grade and 296 from sixth grade. They were drawn from 21 different primary schools. Several criteria were used. To evaluate the internal consistency, Cronbach's alpha was calculated for all summary evaluation criteria (content, coherence, and style); a homogeneity index (H_j) was calculated too. The validity was evaluated by comparing academic levels. As for the results, both reliability and validity indices were high and significant. These results provide empirical evidence for the validity of the summary test.

© 2014 Colegio Oficial de Psicólogos de Madrid. Production by Elsevier España, S.L.U. All rights reserved.

Actualmente, la lectura y la escritura resultan instrumentos de vital importancia para adquirir conocimientos, para acceder a la cultura y para participar en ella. Este papel tan vital nos debe llevar a tomar conciencia de la importancia trascendental que poseen la lectura y la escritura en el mundo educativo, social, cultural y laboral, así como para el propio enriquecimiento personal. La lectura, como la escritura, supone un proceso de formación inseparable del proyecto de una sociedad ilustrada, por lo que constituye lo que se denomina «cultura lectora», esto es, un cómputo de habilidades básicas sobre las que se desarrolla toda una actividad cultural, mediante la cual

las personas nos desenvolvemos y aplicamos nuestros conocimientos y estrategias lectoras/escritas en múltiples contextos de la vida diaria de manera más o menos eficiente. Tanto es así que muchos países desarrollados o en vías de desarrollo están dedicando mucho tiempo y esfuerzo a la investigación sobre este tema y también a su evaluación. Buenos ejemplos son las pruebas internacionales de PISA para estudiantes adolescentes y adultos y las pruebas de TIMMS y PIRLS para niños de 10 años.

Son tres las áreas donde las tareas de lectura y escritura se vienen planteando desde la psicología: desde una perspectiva investigadora, donde se desarrollan análisis y modelos teóricos que sirvan para comprender las variables que implican, desde la perspectiva educativa, donde estos modelos se introducen en el aula como una vía de mejora en estos aprendizajes tan esenciales, o desde una perspectiva

*La correspondencia sobre este artículo debe enviarse a José A. León. Dpto. Psicología Básica. Facultad de Psicología. Universidad Autónoma de Madrid. Campus de Cantoblanco. 29049 Madrid. E-mail: joseantonio.leon@uam.es

clínica, donde la detección precoz de disfunciones o trastornos resultan esenciales para desarrollar lo antes posible una intervención temprana. El trabajo que desarrollamos en este artículo es la primera base para la elaboración de una prueba estandarizada y normalizada de resúmenes para niños de 10 a 12 años, cuya implicación afecta a estas tres perspectivas.

La elección de una prueba sobre resumen responde a varias razones. Una primera es porque no hay aún ninguna prueba baremada sobre resúmenes, siendo una tarea que favorece una metódica exploración de la información almacenada, potencia la estrategia de estructuración de lo leído u oído y fortalece las conexiones entre la nueva información y el conocimiento previo.

Una segunda razón es porque se trata de un buen predictor de la comprensión del lenguaje. Supone planificación y control de la propia comprensión, estimulando habilidades metacognitivas e incrementando, igualmente, la efectividad de los procesos de autorregulación, aspectos que tienen un favorable impacto en la construcción del conocimiento y en el aprendizaje. Existe un importante número de estudios experimentales que han puesto de manifiesto esta relación entre el resumen y la comprensión, así como con el recuerdo del contenido (véase, por ejemplo, Anaya, 2005; Hidi y Anderson, 1986; Kintsch, Steinhart, Stahl y LSA Research Group, 2000; León y Otero, 2009; León y Escudero, en prensa; Rinehart, Stahl y Erickson, 1986) o también la relación establecida entre la validez del resumen como medida de la comprensión (e.g., Al-Shabanah, 2005; Armbruster, Anderson y Ostergat, 1987; Cordero-Ponce, 2000; Jorge y Kreis 2003; Kirkland y Saunders, 1991; León, Escudero, Cañas y Salmerón, 2006; Nelson y Smith, 1992; Taylor, 1983; Thomas y Bridge, 1980; Vadlapudi y Katragadda, 2010; Zipitria, Arruarte, Elorriaga y Díaz de Llaraza, 2007). Así, por ejemplo, Nelson y Smith (1992) evaluaron la calidad de los resúmenes en relación al contenido del pasaje leído, computando el porcentaje de información importante incluida en cada resumen que expresaban como proporción del total de ideas importantes que los evaluadores expertos habían predeterminado. En esta misma línea se sitúan los estudios de Armbruster et al. (1987) o Cordero-Ponce (2000). Además, en diversos estudios aparece recogido que la capacidad de comprensión a edades tempranas es un poderoso predictor de la comprensión en edades más tardías (Oakhill y Cain, 2007). En los últimos cursos de primaria, la comprensión lectora se vuelve especialmente importante, sobre todo porque aporta una de las bases fundamentales para el aprendizaje en secundaria. De hecho, el progreso académico de dichos estudiantes sería muy restringido sin la comprensión lectora y la motivación de leer para aprender (Guthrie et al., 2004).

Una tercera razón obedece a que el resumen posee, además, otras propiedades cognitivas. Así lo han puesto de manifiesto autores como Hadwin, Kirby y Woodhouse (1999), quienes estudiaron la influencia de variables como la memoria de trabajo, la habilidad verbal y el conocimiento previo como predictores de la habilidad de resumir y la eficacia del recuerdo. Los resultados de este estudio indicaron que la calidad de los resúmenes escritos fue el predictor más importante del rendimiento académico y de las diferencias individuales de los estudiantes.

Finalmente, una cuarta razón alude a la relación de la lectoescritura con déficits y trastornos psicológicos. Un importante número de estudios delatan síntomas deficitarios en lectoescritura asociados a trastornos clínicos. En general parece observarse relación entre niños con déficit de atención y/o hiperactividad (TDAH) y esta clase de síntomas. Así lo evidencian, por ejemplo, Buñuel y Olivares (2008), quienes encontraron en niños suecos, daneses y finlandeses puntuaciones elevadas en las escalas de inatención e hiperactividad que correlacionaban con menores habilidades relacionadas con la lectoescritura. Estos autores encontraron además relación entre la presencia de TDAH y un cierto riesgo de deterioro académico, incluso cuando no se cumplían todos los criterios diagnósticos de ese trastorno. Por su parte, Lobo y Lima (2008) realizaron un estudio con niños con y

sin TDAH, detectando que aquellos lectores que presentaban el trastorno obtuvieron un peor desempeño en tareas de lectoescritura.

También se han encontrado evidencias de alteraciones que afectan a procesos ejecutivos como, por ejemplo, a la automaticidad y eficiencia en los procesos de reconocimiento e identificación visual. En tal caso, niños con TDAH que sufren estas alteraciones no son capaces de prestar atención al mismo tiempo a los procesos de comprensión cuando tienen que atender a la codificación y, por ello, se produce una sobrecarga en la capacidad de memoria de trabajo (Sánchez e Hidalgo, 1990), como también una alteración en las funciones ejecutivas, obteniendo inferiores resultados, especialmente en lo relativo a los procesos de planificación (Alvarado, Puente, Jiménez y Arrebillaga, 2011). De igual modo, Dykman y Ackerman (1991; citado en Alvarado et al., 2011) confirmaron que el déficit atencional es el factor esencial que repercute negativamente en niños con TDAH en tareas de comprensión lectora. Existe también alguna evidencia de problemas de aprendizaje relacionados con la escritura en personas que sufren trastorno bipolar, TDAH de tipo combinado, TDA, autismo y espina bífida (Dickerson y Calhoun, 2006). Hay, además, estudios que relacionan estos problemas de escritura con trastornos de internalización y externalización (Willcutt y Pennington, 2000) y con la ansiedad y la depresión (Dickerson y Calhoun, 2006). Ahondando aún más en esta vía, otros autores han encontrado vínculos robustos entre un riesgo incrementado de humor o comportamiento depresivo y problemas de lectura severos y persistentes en niños de entre siete y diez años (Maughan, Rowe, Loeber y Stouthamer-Loeber, 2003).

Todo ello redundando en nuestra idea básica de que para resumir un texto resulta necesario antes comprenderlo, de modo que cuanto mayor sea el nivel de comprensión mejor será la calidad del resumen. En el desarrollo de esta propuesta inicial sobre la tarea de resumen incluimos un texto de primaria, aunque se irán integrando más textos y más niveles académicos. La idea es establecer un criterio general de evaluación que permita evaluar la tarea de resumen y el uso de estrategias de escritura ante diferentes niveles educativos. Se pretende, asimismo, establecer una baremación normalizada en estas edades que permita sospechar, más adelante, déficits y/o trastornos asociados al lenguaje o a los mecanismos de comprensión de cara a un diagnóstico precoz.

Los resúmenes, por tanto, son un tipo de medida que requiere que el lector utilice sus destrezas de producción escrita. No debe resultarnos extraño, por esta misma razón, que aparezcan grandes diferencias individuales en la habilidad de los alumnos para utilizar dichas destrezas. Al realizar un resumen se espera que el lector identifique no solo las ideas principales de lo leído, sino que también sea capaz de reconocer las diferentes conexiones causales dentro de una estructura causal y coherente más global (León, Peñalba y Escudero, 2002; van den Broek, Lynch, Naslund, Levers-Landis y Verduin, 2003). Requiere, en definitiva, no sólo diferenciar entre ideas relevantes y de detalles sino construir el significado en una forma más sucinta y con mayor nivel de abstracción y generalización (van Dijk y Kintsch, 1983).

El supuesto del que partimos es que un buen resumen lleva implícita una buena comprensión, de tal manera que el lector no sería capaz de realizar un buen resumen si antes no comprendiese el texto. Ello implicaría, por ejemplo, establecer una línea argumental en la que la relación de ideas fundamentales expresadas en el texto conllevara un buen nivel de coherencia, una lógica causal y una expresión adecuada. En esta dirección, Jorge y Kreis (2003) generaron un modelo de análisis de los resúmenes para todo tipo de texto cuya calidad era evaluada según cuatro índices: coherencia y cohesión, inclusión de ideas principales, concisión, información sobre la fuente del texto y ausencia de opiniones personales. En este mismo contexto podemos destacar el estudio de Zipitria et al. (2007), quienes desarrollaron un sistema basado en la observación y el análisis del proceso de evaluación del resumen realizado por profesionales expertos. Consiguieron identificar algunas variables como la coherencia, cohe-

sión, adecuación y corrección lingüística y de contenido, que inciden en la calidad del resumen, así como la relevancia relativa de cada una de ellas. Los evaluadores emitían un juicio global del resumen y las puntuaciones parciales se recogían en una escala de 0 a 10. El estudio incluía también análisis estadísticos que pusieron de manifiesto la independencia de las variables. En base a esto, el estudio que nos ocupa pretende evaluar dicha comprensión lectora a través de una tarea de resumen en la que se comprobará en qué medida los alumnos son capaces o no de identificar correctamente las ideas relevantes del texto (contenido), las conexiones causales que las unen (grado de coherencia) y la forma escrita (estilo). De igual modo, el estudio pretende ayudar a la detección de problemas educativos y posibles problemas clínicos en caso de que los niveles de puntuación se posicionen muy por debajo de la media.

Esta prueba ha sido desarrollada dentro de un marco multidimensional que está inspirado en el modelo sobre comprensión lectora desarrollado por León (2004), León, Escudero y Olmos (2012) y León et al. (2009) y también en las pruebas internacionales sobre competencia lectora PISA y PIRLS. Todos ellos coinciden en señalar la estrecha relación existente entre las características del lector, del texto y de las actividades de lectura y escritura.

Objetivos

Se trata de una prueba cuyo objetivo es evaluar la calidad del resumen de los alumnos de primaria. Por un lado permite obtener una evaluación diferenciada de la expresión escrita y por otro detectar diferencias interindividuales como cualquier otra prueba estándar de evaluación.

Método

Participantes

Para la baremación de esta prueba participaron un total de 528 estudiantes de primaria, de los cuales 236 fueron estudiantes de 4º de primaria (9 y 10 años) y 292 a 6º de primaria (11 a 13 años). Todos estos alumnos procedían de 21 centros escolares adscritos a tres comunidades autónomas diferentes: la Comunidad de Madrid (216 estudiantes), la Comunidad de Castilla y León (149 estudiantes) y la Comunidad de Castilla La Mancha (163 estudiantes).

Materiales

Se utilizó un texto narrativo titulado *El hombrecillo sabelotodo* (542 palabras), en formato de diálogo, con una estructura narrativa y lingüística que se acomoda a los alumnos de primaria, extraído de *El patio de mi casa. Realidad y lectura I*. Madrid: Santillana, 1982.

Procedimiento

Esta prueba se diseñó para evaluar a alumnos de 4º y 6º de educación primaria. No obstante, también podría aplicarse a alumnos de otros cursos adyacentes. La aplicación puede realizarse de forma individual o colectiva y la duración es aproximadamente de 15-20 minutos entre la lectura y el resumen solicitado por escrito.

Criterios de evaluación

El objetivo principal de esta propuesta es tratar de generar un mismo tipo de evaluación, un armazón común, que sea válido para cualquier resumen. Para ello, contamos con tres criterios para evaluar el resumen: contenido, coherencia y expresión escrita.

- **Contenido.** El resumen debe reflejar los contenidos esenciales sobre los que se sustenta el texto original. La evaluación de este apartado se basa fundamentalmente en la exposición de las

ideas más relevantes del texto. El contenido se evalúa de 0 a 4 puntos, teniendo en cuenta los siguientes apartados: a) introducción, donde se valora la introducción del personaje y el lugar donde transcurre la historia (hasta 1 punto), b) proceso, en el que se describe el plan en el que se han expuesto las ideas del primer apartado (hasta 1 punto) y c) finalidad, en el que el resumen proporciona una visión general del texto enfatizando las relaciones entre las ideas principales, recogiendo la parte importante de las ideas del texto, la intención del autor y la finalidad o pretensión del texto (hasta 2 puntos).

- **Coherencia.** La coherencia supone la formación de representaciones mentales a partir de la correcta interpretación del argumento del texto. Siguiendo este planteamiento, podríamos identificar resúmenes como estilísticamente pobres (o inclusive incorrectos desde el punto de vista léxico o sintáctico) pero que posean un alto nivel de coherencia. El uso de la correferencia, las relaciones causales, el uso de conectivas y, en general, todo lo que contribuye a una mejora de la comprensión y que aporta el lector se evalúa en este apartado. Se valora de 0 a 4 puntos, teniendo en cuenta los siguientes apartados: a) establecimiento de la red causal, donde se explicita las relaciones causales, el por qué o la consecuencia (hasta 2 puntos), b) orden de las ideas, que refleja el orden secuencial de las ideas, con cierta lógica o cronología, no habiendo saltos grandes entre ellas (hasta 1 punto) y c) cohesión escrita, donde se evalúan la claridad y la cohesión entre las ideas explicitadas en el resumen (hasta 1 punto).
- **Expresión escrita.** Hace referencia al estilo y forma, a la personalización y a la síntesis adecuada (0 a 2 puntos). Se valoran dos aspectos en este apartado. Por un lado, si el resumen es *personalizado*, esto es, se elabora sobre una base original, si el contenido se expresa con las propias palabras del que resume. Se evalúa aquí el uso de paráfrasis y se penaliza o no se puntúa si sigue la estrategia de suprimir y copiar (hasta 1 punto). Por otro lado, se evalúa la *expresión sintética o concisa*, en cuyo caso se valora la síntesis, si no hay frases innecesarias, superficiales o ideas repetidas (hasta 1 punto).

Resultados

Estadísticos descriptivos

En la tabla 1 se muestran los estadísticos descriptivos de las puntuaciones del resumen del texto. En dicha tabla se muestran la media, la desviación típica y el número de casos para 4º y 6º de primaria. Se presentan los descriptivos de la evaluación del contenido, la coherencia y el estilo.

Análisis de la consistencia interna

En primer lugar, se calcula el alfa de Cronbach utilizando los ocho criterios que se emplearon para evaluar la calidad de los resúmenes. Estos criterios son:

- **Contenido:** (1) personajes y lugar, cuya puntuación tiene un rango de 0 a 1 puntos, (2) proceso, con una puntuación dicotómica de 0 (no lo contiene el resumen) y 1 (lo contiene), y (3) finalidad, también con una puntuación de 0 a 2.
- **Coherencia:** (4) relación causal, cuya puntuación oscila entre 0 y 2 puntos, (5) orden de las ideas, con una puntuación dicotómica 0 y 1 y (6) forma de escritura, también con una puntuación dicotómica 0 y 1.
- **Estilo:** (7) personalización, con una puntuación dicotómica 0 y 1 y (8) expresión sintética, también con una puntuación 0 y 1.

Todo ello hace que los valores de los resúmenes oscilen entre un mínimo de 0 y un máximo de 10 puntos.

En segundo lugar, se presentan los resultados para los ocho criterios individuales para, de esta forma, poder valorar la contribución

Tabla 1
Media, desviación típica y *N* de los resúmenes del texto "El hombrecillo"

		Contenido	Coherencia	Estilo	Resumen: puntuación total	
		Media (DT)	Media (DT)	Media (DT)	Media (DT)	<i>N</i>
Curso académico	Cuarto de primaria	2.05 (1.26)	2.00 (1.15)	1.52 (0.75)	5.57 (2.81)	232
	Sexto de primaria	2.88 (1.22)	2.78 (1.18)	1.85 (0.40)	7.51 (2.57)	289
	Total	2.51 (1.30)	2.43 (1.23)	1.70 (0.60)	6.65 (2.84)	521

de cada criterio al alfa de Cronbach, sus medias o índice de dificultad, sus desviaciones típicas y el índice de homogeneidad (H_j):

- **Índice de dificultad del elemento:** este índice refleja la dificultad de los elementos en la muestra de tipificación. Se calcula a partir de la proporción de aciertos en cada ítem. Valores próximos a 1 indican que el ítem es muy fácil y valores próximos a 0 indican que el ítem es muy difícil.
- **Desviación típica:** desviación típica de las puntuaciones en el ítem de los sujetos de la muestra de tipificación.
- **Índice de homogeneidad del elemento:** se calcula mediante la correlación corregida del elemento con el total de la escala. Indica hasta qué punto el elemento está en consonancia con lo que mide la escala.

Respecto al estudio de la consistencia interna del texto tenemos que el alfa de Cronbach también es muy alto, tal y como se muestra en la tabla 2, ya que para el conjunto total de la muestra éste es de .86, para la muestra de cuarto de primaria es también .86 y para la de sexto su valor es de .84.

Cabe destacar aquí que los índices de homogeneidad, H_j , son altos y positivos en los ocho criterios evaluados, contribuyendo a aumentar todos ellos en su conjunto la fiabilidad de la evaluación del resumen de este texto. El aspecto evaluado que resultó más difícil en este

Tabla 2
Índice de consistencia interna alfa de Cronbach para cuarto y sexto de primaria en el texto "El hombrecillo"

Curso	Alfa de Cronbach
Cuarto de primaria	.86
Sexto de primaria	.84
Total	.86

Tabla 3
Índices psicométricos de los criterios evaluados para cuarto y sexto de primaria en el texto "El hombrecillo"

Curso	Cuarto de primaria			Sexto de primaria		
	Media	DT	H_j	Media	DT	H_j
Personaje y lugar	0.82	0.27	0.48	0.91	0.22	.55
Proceso	0.53	0.42	0.76	0.74	0.35	.73
Finalidad	0.71	0.80	0.70	1.24	0.82	.80
Relación causal	0.72	0.67	0.82	1.22	0.73	.84
Orden de ideas	0.57	0.39	0.73	0.77	0.36	.83
Forma de escritura	0.72	0.38	0.27	0.81	0.34	.30
Personalización	0.80	0.35	0.67	0.95	0.19	.52
Expresión sintética	0.71	0.44	0.75	0.91	0.26	-.65

Tabla 4
Fiabilidades interjueces (Pearson) para el contenido, coherencia y del total de los resúmenes del texto "El hombrecillo"

	Contenido			Coherencia			Total		
	Juez1	Juez2	Juez3	Juez1	Juez2	Juez3	Juez1	Juez2	Juez3
Juez1	.85	.97	.94	.73	.85	.95	.79	.94	.97
Juez2		.85	.73		.69	.75		.79	.76
Juez3			.93			.84			.91

texto fue el *proceso* en 4º de primaria (.53), siendo mejor en 6º de primaria (media 0.74), por lo que una de las mayores diferencias entre los resúmenes de ambos cursos se encuentra precisamente en este criterio (véase tabla 3).

Estudio de la fiabilidad interjueces

Para analizar la fiabilidad del texto se contó con cuatro evaluadores que calificaron 20 resúmenes escogidos al azar. La fiabilidad interjueces en cuanto al contenido medido por la correlación de Pearson osciló entre .69 y .97, siendo por tanto la fiabilidad sumamente aceptable (véase tabla 4). El coeficiente de correlación intraclase para evaluar la fiabilidad que tendría un evaluador (un profesor, por ejemplo) cualquiera fue de .86 en el contenido, .73 en la coherencia y de .82 en el total. Para el cálculo se asumió un modelo de efectos aleatorios tanto para los profesores como para los estudiantes (por lo tanto, la situación más generalizable de todas).

Estudio de validez : comparación entre niveles académicos

Este análisis lo centramos en las comparaciones de medias entre los diferentes cursos académicos analizados. Comparamos los grupos de 4º y 6º en el rendimiento que obtuvieron en los distintos apartados de las evaluaciones de cada resumen mediante una *t* de Student. Se detectaron diferencias significativas a favor de 6º de primaria en todos los apartados de evaluación del texto. En contenido, la media para 4º de primaria fue de 2.05 y para 6º de 2.88, $T(519) = 7.66$, $p < .001$, $\delta = .68$. En coherencia, la media para 4º de primaria fue de 2.00, mientras que para 6º fue de 2.78, $T(519) = 7.50$, $p < .001$, $\delta = .66$. En estilo, la media para 4º fue de 1.52 y para 6º de 1.86, $T(519) = 6.60$, $p < .001$, $\delta = .58$. Para el total de resumen, la media para 4º fue de 5.57 y para 6º de 7.51, $T(519) = 8.23$, $p < .001$, $\delta = .73$.

Baremación

En las tablas de baremos que aparecen en este apartado se muestran tres tipos de puntuaciones, como puede observarse en las tablas 5, 6 y 7. La que aparece en la columna de la izquierda bajo el encabezado "PD" de cualquiera de las tablas baremo es la puntuación directa (PD). Esta se refiere a la suma que un resumen obtiene tras aplicar todos los criterios de evaluación al mismo. Una puntuación directa tiene el inconveniente de que no aporta información sobre dónde se sitúa un estudiante en relación a su grupo de referencia, por lo que necesitamos otro tipo de puntuaciones que en ese sentido sean más informativas; por eso, las tablas de baremos incluyen otros dos tipos de puntuaciones más. En la columna de la derecha aparece el encabezado "percentil". Una puntuación percentil tiene la virtud de que informa del porcentaje de estudiantes del grupo de referencia que obtiene una puntuación directa dada como esa o más baja en el resumen. Así, un percentil 95 asociada a una puntuación directa de 9, significa que el

95% de los estudiantes de un nivel académico determinado obtienen una puntuación de 9 o menos en el resumen. Por lo tanto, nos sitúan al estudiante en relación a su grupo de referencia. Por último, también se ofrece una puntuación Z bajo el encabezado "puntuación Z". Esta puntuación informa del número de desviaciones típicas que se aleja un estudiante de la media de su grupo de referencia. Así, una puntuación Z de 1.5 significa que ese estudiante está 1.5 desviaciones típicas por encima de la media de su grupo de referencia.

Consideraciones finales

Esta prueba fue construida para responder a la necesidad de medir la calidad de los resúmenes. Los resultados obtenidos indican que la prueba RESUMeV es un instrumento válido para dichos resúmenes en niños de 9 a 12 años. Las evidencias de fiabilidad y validez obtenidas garantizan que la prueba representa adecuadamente al constructo que trata de medir y que las interpretaciones propuestas se

Tabla 5
Puntuación total en el resumen del texto "El hombrecillo"

Cuarto de primaria			Sexto de primaria		
PD	Puntuación Z	Percentil	PD	Puntuación Z	Percentil
0	-1.98	1	0	-2.92	1
0.5	-1.81	3	0.5	-2.73	1
1	-1.63	6	1	-2.53	2
1.5	-1.45	11	1.5	-2.34	3
2	-1.27	16	2	-2.14	4
2.5	-1.09	24	2.5	-1.95	7
3	-0.92	27	3	-1.76	9
3.5	-0.74	30	3.5	-1.56	11
4	-0.56	34	4	-1.37	15
4.5	-0.38	40	4.5	-1.17	17
5	-0.2	43	5	-0.98	20
5.5	-0.02	49	5.5	-0.78	24
6	0.15	54	6	-0.59	27
6.5	0.33	61	6.5	-0.39	32
7	0.51	67	7	-0.2	36
7.5	0.69	75	7.5	0	41
8	0.87	81	8	0.19	53
8.5	1.04	85	8.5	0.39	57
9	1.22	88	9	0.58	64
9.5	1.4	92	9.5	0.77	73
10	1.58	100	10	0.97	100

Tabla 6
Puntuación contenido en el resumen del texto "El hombrecillo"

Cuarto de primaria			Sexto de primaria		
PD	Puntuación Z	Percentil	PD	Puntuación Z	Percentil
0	-1.62	3	0	-2.37	2
0.5	-1.23	18	0.5	-1.96	7
1	-0.83	34	1	-1.55	15
1.5	-0.43	48	1.5	-1.14	22
2	-0.04	58	2	-0.73	30
2.5	0.36	71	2.5	-0.31	37
3	0.76	78	3	0.10	52
3.5	1.15	81	3.5	0.51	59
4	1.55	100	4	0.92	100

Tabla 7
Puntuación coherencia + estilo en el resumen del texto " El hombrecillo"

Cuarto de primaria			Sexto de primaria		
PD	Puntuación Z	Percentil	PD	Puntuación Z	Percentil
0	-2.04	3	0	-2.68	1
0.5	-1.75	7	0.5	-2.39	2
1	-1.46	15	1	-2.10	3
1.5	-1.17	22	1.5	-1.81	7
2	-0.88	27	2	-1.53	9
2.5	-0.59	33	2.5	-1.24	13
3	-0.30	38	3	-0.95	15
3.5	-0.01	47	3.5	-0.66	20
4	0.28	58	4	-0.37	33
4.5	0.57	72	4.5	-0.08	44
5	0.86	86	5	0.21	61
5.5	1.15	89	5.5	0.50	66
6	1.44	100	6	0.79	100

ajustan a las puntuaciones obtenidas. Como instrumento válido y fiable no solo permite detectar qué estudiantes están desempeñando de manera más pobre estos procesos, sino también evaluar los cambios en estos procesos que como es sabido se corresponden con otros muchos problemas asociados. Como se ha señalado anteriormente, mediante la evaluación de esta tarea de resumen, podrá comprobarse en qué medida los alumnos son o no capaces de identificar correctamente las ideas relevantes del texto (contenido), las conexiones causales que las unen (grado de coherencia) y la forma escrita (estilo). Pero también unas bajas puntuaciones en esta prueba pueden inducirnos a sospechar la existencia de algunos problemas educativos o clínicos. Como se recordará, un importante número de estudios delatan síntomas deficitarios en lectoescritura asociados a trastornos clínicos, como los asociados a niños con déficit de atención y/o hiperactividad (TDAH). Todo ello puede resultar de utilidad a la hora de elaborar un diagnóstico y de diseñar su intervención.

Extended Summary

No one would argue that nowadays being able to read is of vital importance. Written discourse continues to be the main means for the organized transmission of knowledge and written materials pervade every aspect of our lives, be it educational, social, cultural, professional, or for our own personal improvement. Reading is of such transcendental importance that many developed and developing nations are expending considerable resources on reading research. One example of this is the PISA Project which, since the year 2000, has attempted to establish common criteria among OCDE countries for the evaluation of reading competence as well as to inform future education policies. The evaluation of reading and the identification of its component processes can thus be used to predict academic performance and establish appropriate levels of skill for adult life.

One area of text comprehension research that has most interested psychologists and discourse researchers concerns the processes that occur during the comprehension and summarizing phases of reading. Comprehension and summarizing are very closely related. In fact, some researchers have suggested that if readers are not able to summarize a passage, then they have not understood it. A generally acknowledged practice consists of using a summary to organize and emphasize the most relevant content of the text. Although the summary concept is imprecise, summaries themselves hold a significant place in scientific texts, and their effectiveness in improving compre-

hension and recall is generally recognized. When readers summarize a passage, they tend to form a nucleus of information, a core concept that represents a general vision of the text in a coherent way. Synthesis and coherence are two key aspects of a good summary. In order to summarize a text, a reader must read and comprehend the material, isolate the main ideas, and convey those ideas succinctly. In general, we can assume that a summary is a concise statement of the most important information in a text. A summary should describe most of the main ideas (or main topics) in the text.

Some authors suggested that summaries should be evaluated on the basis of four criteria. Jorge and Kreis (2003), in their summary analysis model which draws on the work of several writers, used five parameters to measure the quality of summaries: cohesion and coherence, inclusion of the main ideas contained in the source text, conciseness, information about the source text, and absence of personal opinion. A similar study was carried out by Zipitria, Arruarte, Elorriaga, and Díaz de Llaraza (2007) who developed a system based on observing and analysing the processes used by expert evaluators of summaries. Zipitria and his colleagues identified variables (e.g., coherence and cohesion, appropriate and correct language, appropriate, correct and relevant content) and examined the overall contribution each of these made to the quality of the summary. The evaluators gave a global mark to the summary as well as marking each variable on a 0 to 10 scale. This study also reported a statistical analysis, which showed a certain independence of the variables. A generalized conclusion of all these studies is that a summary task encourages deep understanding of the text because it requires active construction of the meaning as opposed to merely choosing one response from several alternatives or answering isolated questions. As such, writing summaries can be a valuable complementary instrument to traditional methods (i.e., multiple choice) of evaluating students' comprehension of science texts.

We proposed that summaries should be evaluated on the basis of three criteria which were found to be appropriate for narrative texts (León, Escudero, & Olmos, 2012). The three criteria were *content*, *coherence* and *cohesion* or *written expression*. *Content* concerns the extent to which the summary reflects the essential content of the text. Evaluation is based principally on the textbase, that is, on whether the most relevant ideas have been included. *Coherence* is the main characteristic of a reader's mental representation of text content. Coherence relations are constructed in the reader's mind and depend on the skills and knowledge that the reader brings to the situation.

This analysis requires differing integrated levels of representation, including text-based models (based on topics and ideas from the text) and situational models (based on the reader's prior knowledge). As a result, summarizing is a highly effective means of constructing and integrating new knowledge as well as connecting causal relations between the relevant ideas, including reasons and consequences. These aspects should be clear and explicit in a good summary. Along with other details supplied by the reader, the aspects also give greater coherence to the summary. Finally, *cohesion* or *written expression* refers to the style and form of the summary. The characteristics that are evaluated positively are the use of paraphrasing, correct synthesis, and whether the summary is 'personalized', that is, whether the writer has used his or her own words. Characteristics that are evaluated negatively or even penalized are the inclusion of unnecessary or irrelevant information, the repetition of ideas, the use of the copy-paste strategy, and whether the summary is too long (and, hence, it probably contains superfluous details).

The purpose of this work was to collect construct and criterion-related evidence of validity for a summary test (RESUMeV) designed to assess students from second and fourth grade. The sample of this study consisted of 528 children, 236 from fourth grade, and 296 from sixth grade. They were taken of 21 different primary schools.

Several criteria were conducted. To evaluate the internal consistency, Cronbach's alpha was calculated for all summary evaluation criteria (content, coherence and style) and its homogeneity index (Hj) was calculated too. Cronbach's alpha was very high, reaching .86 for the total sample and fourth grade, and .84 for sixth grade. The validity was evaluated with the comparison between academic levels. The results that were obtained both in reliability and validity criteria were high and significant. The present results provide empirical evidence for the validity of the summary test.

Each summary was graded by four judges who had been trained in the evaluation criteria described above in the Methods section. The results of the Pearson correlation coefficient suggest that interrater reliability in the marking of the text was high (from $r = .69$ to $r = .93$). The intraclass correlation coefficient to test the reliability that would have any of the evaluators was .86 for the content, .73 for the coherence, and .82 for the total. To calculate it, a model of random effects was assumed for both, professors and students.

Therefore, summary writing can be said to have a predictive value for reading comprehension and, in turn, of reading competence. The three criteria (i.e., content, coherence, and cohesion) are relevant to evaluating summary tasks adequately. Finally, we highlight that the summary writing is a valuable learning activity that requires more research in order to check or suspect clinical problems or deficits in learning. This requires more research in the near future.

Conflicto de intereses

Los autores de este artículo declaran que no tienen ningún conflicto de intereses.

Financiación

Este trabajo ha sido subvencionado por el MINNECO PSI2013 47219-P.

Referencias

- Al-Shabanah, A. I. (2005). *Summarization Strategies Adopted by the Senior Level Female Students of the Department of English at King Saud University* (A Thesis Submitted in Partial Fulfillment of the Requirements for the Master's Degree in Applied Linguistics in the Department of English at the College of Arts). King Saud University.
- Alvarado, J. M., Puente, A., Jiménez, V. y Arrebillaga, L. (2011). Evaluating Reading and Metacognitive Deficits in Children and Adolescents with Attention Deficit Hyperactivity Disorder. *The Spanish Journal of Psychology*, 14, 62-73.
- Anaya, D. (2005). Efectos del resumen sobre la mejora de la metacompreensión, la comprensión lectora y el rendimiento académico. *Revista de Educación*, 337, 281-294.
- Armbruster, B. B., Anderson, T. H. y Ostertag, J. (1987). Does text structure/ summarization instruction facilitate learning from expository text? *Reading Research Quarterly*, 22, 331-346.
- Buñuel, J. C. y Olivares, M. (2008). Puntuaciones elevadas en uno o varios de los síntomas nucleares del TDAH en niños sin el trastorno podrían asociarse a una disminución en el rendimiento de algunas habilidades académicas. *Evidencias en Pediatría*, 4(1), 4-17.
- Cordero-Ponce, W. L. (2000). Summarization instruction: Effects on foreign language comprehension and summarization of expository texts. *Reading Research and Instruction*, 39, 329-350.
- Dickerson, S. y Calhoun, S. L. (2006). Frequency of reading, math and writing disabilities in children with clinical disorders. *Learning and Individual Differences*, 16, 145-157.
- Dykman, R. A., y Ackerman, P. I. (1991). Attention deficit disorder and specific reading disability. Separate but often overlapping disorder. *Journal of Learning Disabilities*, 24, 96-102.
- Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., ... Tonks, S. (2004). Increasing Reading Comprehension and Engagement Through Concept-Oriented Reading Instruction. *Journal of Educational Psychology*, 96, 403-423.
- Hadwin, A. F., Kirby, J. R. y Woodhouse, R. A. (1999). Individual differences in note-taking, summarization and learning from lectures. *Alberta Journal of Educational Research*, 45, 1-17.
- Hidi, S. y Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research*, 56, 473-493.
- Jorge, R. y Kreis, L. (2003). The influence of Reading upon writing in EFL student's summarizing process - An experiment. *Fragmentos*, 25, 103-120.
- Kintsch, E., Steinhart, D., Stahl, G. y LSA Research Group (2000). Developing summarization skills through the use of LSA-based feedback. *Interactive Learning Environments*, 8, 87-109.
- Kirkland, M. R. y Saunders, A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly*, 25, 105-121.
- León, J. A. (2004). *Adquisición de Conocimiento y comprensión: Origen, evolución y método*. Madrid: Biblioteca Nueva.
- León, J. A. y Escudero, I. (en prensa). Understanding Causality in Science Discourse for Middle and High School Students. Summary task as a Strategy for Improving Comprehension. In K. L. Santi y D. Reed (Eds.), *Improving Comprehension for Middle and High School Students*. Springer.
- León, J. A., Escudero, I. y Olmos, R. (2012). *Ecomplec: Una propuesta de evaluación de la comprensión lectora en Primaria y Secundaria*. Madrid: TEA Ediciones.
- León, J. A., Escudero, I., Olmos, R., Sanz, M.ª, Dávalos, T. y García, T. (2009). Ecomplec: Un modelo de evaluación de la comprensión lectora en diversos tramos de la Educación Secundaria. *Psicología Educativa*, 15, 123-142.
- León, J. A., Olmos, R., Escudero, I., Cañas, J. J. y Salmerón, L. (2006). Assessing Summaries With Human Judgments Procedure and Latent Semantic Analysis in narrative and expository texts. *Behavior Research Methods, Instruments, & Computers Journal*, 38, 616-627.
- León, J. A. y Otero, J. (2009). Procesos metacognitivos y tareas de resumen en la lectura: implicaciones para la interpretación de los resultados del estudio PISA. *Revista AULA*, 179, 47-51.
- León, J. A., Peñalba, G. E., y Escudero, I. (2002). "Profe, ¿puedo preguntar?". Una breve introducción a la interacción de preguntas y respuestas entre profesor y alumno. *Psicología Educativa*, 8, 107-126.
- Lobo, P. A. S. y Lima, L. A. M. (2008). Comparison of children performance in Reading words with and without attention deficit/hyperactivity disorder. *CEFAC*, 10, 471-483.
- Maughan, B., Rowe, R., Loeber, R. y Stouthamer-Loeber, M. (2003). Reading Problems and Depressed Mood. *Journal of Abnormal Child Psychology*, 31, 219-229.
- Nelson, J. R., y Smith, D. J. (1992). The effects of teaching a summary skills strategy to students identified as learning disabled on their comprehension of science text. *Education and Treatment of Children*, 15, 228-243.
- Oakhill, J. y Cain, K. (2007). Issues of causality in children. En D. S. McNamara (Ed.), *Reading comprehension strategies. Theories, interventions and technologies* (pp. 47-71). Nueva York, LEA.
- Rinehart, S. D., Stahl, S. A. y Erickson, G. L. (1986). Some effects of summarization training on reading and studying. *Reading Research Quarterly*, 21, 422-438.
- Sánchez, J. e Hidalgo, M. D. (1990). Implicaciones de la codificación visual en el retraso específico en lectura. *Psicothema*, 2, 35-48.
- Taylor, K. (1983). Can college students summarize? *Journal of Reading*, 26, 524-528.
- Thomas, S. y Bridge, C. A. (1980). A comparison of subjects' cloze scores and their ability to employ macrostructure operations in the generations of summaries. In M. L. Kamil y A. J. Moe (Eds.), *Perspectives on reading research and instruction. Twentieth yearbook of the national reading conference* (pp. 69-77). Washington, DC.
- Vadlapudi, R. y Katragadda, R. (2010). Quantitative evaluation of grammaticality of summaries. *Computational Linguistics and Intelligent Text Processing*, 64, 736-747.
- Van den Broek, P., Lynch, J. S., Naslund, J., Ievers-Landis, C. E. y Verduin, K. (2003). The development of comprehension of main ideas in narratives: Evidence from the selection of titles. *Journal of Educational Psychology*, 95, 707-718.
- Van Dijk, T. A., Kintsch, W. y Van Dijk, T. A. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Willcutt, E. G. y Pennington, B. F. (2000). Psychiatric Comorbidity in Children and Adolescents with Reading Disability. *Journal of Child Psychology and Psychiatry*, 41, 1039-1048.
- Zipitria, I., Arruarte, A., Elorriaga, J. A. y Díaz de Llaraza, A. (2007). Hacia la automatización de la evaluación de resúmenes desde la experiencia cognitiva. *Revista Iberoamericana de Informática Educativa*, 5, 49-61.