



Evaluación de las estrategias y procesos de comprensión: el Test de Procesos de Comprensión

Tomás Martínez, Eduardo Vidal-Abarca, Pilar Sellés & Ramiro Gilabert

To cite this article: Tomás Martínez, Eduardo Vidal-Abarca, Pilar Sellés & Ramiro Gilabert (2008) Evaluación de las estrategias y procesos de comprensión: el Test de Procesos de Comprensión, *Infancia y Aprendizaje*, 31:3, 319-332, DOI: [10.1174/021037008785702956](https://doi.org/10.1174/021037008785702956)

To link to this article: <https://doi.org/10.1174/021037008785702956>



Published online: 23 Jan 2014.



Submit your article to this journal [↗](#)



Article views: 252



View related articles [↗](#)



Citing articles: 15 View citing articles [↗](#)

Evaluación de las estrategias y procesos de comprensión: el Test de Procesos de Comprensión

TOMÁS MARTÍNEZ, EDUARDO VIDAL-ABARCA, PILAR SELLÉS
Y RAMIRO GILABERT

Universitat de València



Resumen

En este trabajo se presenta un nuevo test de evaluación de la comprensión lectora llamado Test de Procesos de Comprensión (TPC) que, a diferencia de la mayoría de los test en español actuales, está basado en un análisis de procesos de comprensión contemplados en la mayoría de los modelos cognitivos actuales. El test es aplicable a escolares entre 11 y 16 años, de fácil aplicación y corrección. Los resultados empíricos muestran que el test cumple con los requisitos psicométricos estándar de homogeneidad, validez y fiabilidad, y muestra un alto poder para discriminar entre estudiantes de diferentes niveles escolares. Supone un avance metodológico en la elaboración de pruebas de evaluación fundamentadas en modelos cognitivos ampliamente aceptados.

Palabras clave: Test de evaluación, comprensión lectora.

Evaluation of comprehension strategies and processes: Test of Comprehension Processes

Abstract

A new test to assess reading comprehension skills called Test of Comprehension Processes (TCP) is presented. Contrary to other reading comprehension test published in Spanish, TCP is based on an analysis of the reading comprehension processes commonly accepted by current reading comprehension theories. The test can be applied to children between 11 and 16 years of age. It is easy to apply and score. Empirical results show that TCP is reliable, valid and can discriminate among children of different school levels. The development of TCP shows a new way to develop school test grounded on broadly accepted cognitive models.

Keywords: Evaluation test, reading comprehension.

Agradecimientos: Este proyecto ha sido financiado por el Instituto Calasanz de Ciencias de la Educación ICCE y por el Ministerio de Educación y Ciencia: Plan Nacional de Investigación Científica 2006-08 (SEJ2005-04500).

Correspondencia con los autores: Departamento de Psicología Evolutiva y de la Educación. Universitat de València. Av. Blasco Ibáñez, 21. 46010 Valencia. Teléfono: 96386447. Fax: 963864671. Email: martineg@uv.es - vidala@uv.es - senopi@uv.es - gilabert@uv.es

La historia de la evaluación de la comprensión lectora es larga y ha estado jalonada por ciclos de gran esperanza y altas expectativas alternados con periodos de frustración y desencuentro (Pearson y Hamm, 2005). Últimamente hay un aumento del interés por la evaluación de la comprensión, quizás por influencia de estudios internacionales como PISA (*Program for International Students Assessment*) promovido por la OCDE (OECD, 2002, 2005), dirigidos a evaluar la calidad de los diferentes sistemas educativos, en función de habilidades básicas entre las que se incluye la lectura.

En el contexto español, un problema importante es que la mayoría de los tests de comprensión no han sido elaborados desde una base teórica explícita que respalde el diseño de ítems y textos. Una excepción son los tests PROLEC (Cuetos, Rodríguez y Ruano, 1996) y PROLEC-SE (Ramos y Cuetos, 1999) que parten de un modelo cognitivo procesual claro, aunque, en lo que respecta a la evaluación de la comprensión, la distinción que establecen entre procesos de comprensión literal e inferencial, y consecuentemente entre preguntas literales e inferenciales, resulta demasiado genérica si tenemos en cuenta los avances producidos en los últimos años en la investigación sobre procesos de comprensión. Así, en la mayoría de los test de comprensión en español hay gran distancia entre conocimientos teóricos y procedimientos de evaluación. Ésta es la principal motivación para diseñar el Test de Procesos de Comprensión (TPC) que se describe en este trabajo, el cual se fundamenta en los avances de comprensión realizados en los últimos años. El marco teórico global que subyace a nuestro test es el Modelo Construcción-Integración (C-I) de Kintsch (1988, 1998), sin duda el modelo simbólico de comprensión de espectro más amplio y más citado en la literatura científica, completado con aportaciones del modelo Constructivista (Graesser, Singer y Trabasso, 1994). En el modelo C-I la comprensión es definida como un proceso cíclico de construcción e integración de información. El modelo distingue dos niveles de representación: la base del texto y el modelo de la situación. La *base del texto* estaría delimitada por las operaciones relacionadas con la extracción del significado a partir de la información esencial, las relaciones semánticas o retóricas entre las ideas del texto y la detección de los posibles problemas de cohesión o coherencia. En el nivel del *modelo de la situación* el lector iría más allá del propio texto, puesto que incorpora sus conocimientos previos, las inferencias sobre las intenciones del autor, el estilo utilizado, la resolución de los problemas de coherencia y, en última instancia, reformula una interpretación del texto.

El modelo Constructivista de Graesser *et al.* (1994) aporta una distinción interesante tanto desde el punto de vista teórico como aplicado, entre inferencias que conectan-elementos-textuales (*i. e., text-connecting inferences*) e inferencias basadas-en-el-conocimiento (*i. e., knowledge-based inferences*). Un caso típico de inferencias que conectan-elementos-textuales son las inferencias anafóricas en las que el lector debe encontrar un antecedente textual para conectar otro elemento de información textual. Por su parte, las inferencias basadas-en-el-conocimiento son aquellas en las que el lector ha de incorporar estructuras de conocimiento previo para entender una determinada situación expuesta en un texto. Desde el punto de vista teórico la distinción es importante porque la fuente de las inferencias es diferente en cada caso. Así, en las inferencias que conectan-elementos-textuales la fuente primordial es el propio texto, que induce al lector a hacer la inferencia, sin embargo, en el caso de las inferencias-basadas-en-el-conocimiento la fuente principal son las estructuras de conocimiento del lector. Ésta distinción también es importante porque mientras que el primer tipo de inferencias cuenta con marcadores textuales, que pueden servir de aviso al lector, y cuyo uso puede ser enseñado mediante reglas claras (Graesser, en prensa), las inferencias basadas-en-el-conocimiento suelen carecer de marcadores textuales. En este caso es el lec-

tor quien debe hacer un esfuerzo, especialmente costoso en el caso de textos expositivos donde las inferencias son menos frecuentes (Graesser y Goodman, 1985) y más difíciles de elaborar (Coté, Goldman y Saul, 1998), para ir relacionando unas ideas con otras a fin de lograr formar un buen modelo de la situación.

La comprensión desde este punto de vista sería, por tanto, el producto de una serie de procesos cognitivos básicos que implican: (a) formar y conectar proposiciones dentro de una frase (Kintsch y Keenan, 1973; Ratcliff y McKoon, 1978), (b) realizar inferencias tanto las que conectan elementos textuales como las basadas en el conocimiento (Long, Oppy y Seely, 1994; Magliano, Trabasso y Graesser, 1999; Oakhill, Cain y Yuill, 1998) y (c) formar macroideas o estructuras de integración de la información (López y Alonso-Tapia, 1996; van den Broek, Lynch, Naslund, Ievers-Landis y Verduin, 2003). Todos estos procesos se llevan a cabo en la memoria de trabajo (Daneman y Carpenter, 1980; Kintsch, 1998), en una secuencia que va construyendo una representación coherente del texto en la que se incluye tanto la información extraída del texto como las inferencias generadas por el lector. A partir de este marco teórico, el objetivo de nuestro trabajo se concreta en la elaboración de una escala de comprensión, capaz de evaluar adecuadamente estos procesos cognitivos y el grado en que el lector es capaz de construir una representación interna del texto.

Como decíamos más arriba, la mayoría de los tests de comprensión en español carecen de un modelo teórico de comprensión, lo que les lleva a referirse a capacidades genéricas de comprensión sin una articulación suficiente entre las capacidades y los ítems que supuestamente miden las primeras. A continuación analizamos brevemente algunos de los tests españoles de comprensión más utilizados para ilustrar esta laguna.

En cuanto a la comprensión lectora, la *Escala Magallanes de Lecto-escritura TALE 2000* (Urío, Toró y Cervera, 2000) utiliza tres textos narrativos diferentes según el nivel de los estudiantes. Tras la lectura el lector debe responder preguntas de alternativa múltiple en las que debe identificar los elementos relevantes de una narración. El test únicamente menciona habilidades generales de razonamiento, atencionales, conocimiento de vocabulario y conocimiento morfosintáctico para justificar la elaboración de los ítems.

El test ECL (1y 2) de *Evaluación de la Comprensión Lectora* (de la Cruz, 1997) emplea diferentes tipos de textos narrativos y expositivos, graduados según la edad de los sujetos. Las preguntas son de elección múltiple y se controla el tiempo de respuesta de los niños. Este test se basa en la medida de la cantidad de información que el estudiante es capaz de extraer de un texto. Al referirse a conocimientos y capacidades que el lector ha de utilizar menciona las siguientes: conocimiento de las palabras, seguimiento de instrucciones sencillas, conocimiento de antónimos y sinónimos, retención y evocación de la información, integración del significado de frases, inferencia del significado de palabras no conocidas y habilidad para captar la idea principal. El test, sin embargo, no efectúa una clasificación de las preguntas a partir de esas capacidades generales, ni explica cómo desde éstas se han elaborado los ítems.

La *Prueba de Comprensión Lectora* (Lázaro, 1999) utiliza 18 textos breves con diversas estructuras, a los que siguen una serie de preguntas de alternativa múltiple y de relacionar pares de expresiones. Las preguntas se refieren a aspectos generales tales como las inferencias basadas en las características del texto, la intención del autor o del mensaje, la forma y el estilo de las expresiones, el ritmo y la métrica, y la percepción lectora.

La prueba de *Comprensión Lectora, Cloze Test, CLT* (Suárez y Meara, 1982) se compone de dos partes A y B. La primera contiene 4 textos (diálogo, descripción, instrucciones de juego y pasaje narrativo) y la segunda se compone única-

mente de un texto en el que se mezclan la descripción, la narración y el diálogo. Las tareas que proponen a los sujetos son de tipo "cloze", en las que se deben rellenar una serie de huecos que hay en los textos. Los ítems de esta prueba evalúan el vocabulario, la sensibilidad a la estructura del texto y a los condicionamientos sintácticos y semánticos que pueden influir en su comprensión.

En ninguna de las pruebas mencionadas encontramos una mención explícita a un modelo teórico que inspire la prueba, ni se explicitan a los criterios de elaboración de ítems para medir procesos de comprensión o la validación de los mismos. Como dijimos, un caso aparte son el *PROLEC* (Cuetos *et al.*, 1996) y el *PROLEC-SE* (Ramos y Cuetos, 1999), ambos fundamentados en un modelo teórico procesual. En las subpruebas de comprensión los estudiantes deben leer unos textos y responder posteriormente a preguntas abiertas, entre las que se distinguen preguntas literales e inferenciales. La distinción supone un claro avance respecto al resto de pruebas, pero resulta demasiado genérica, aspecto justificado porque esta prueba evalúa todos los procesos de la lectura y no sólo los procesos de comprensión.

Además de la cuestión central de un modelo teórico que fundamente a un test de comprensión, debemos subrayar otras tres cuestiones específicas que merecen consideración adicional en la elaboración de una prueba de este tipo. La primera es la relación entre el test y las situaciones de enseñanza y aprendizaje, lo que determina la elección del tipo de textos y, por ende, de las preguntas. Algunos de los tests mencionados incluyen gran variedad de textos, es decir, narrativos y expositivos (p. ej., *PROLEC*, *ECL*, *Prueba de Comprensión Lectora*). Alguno incluye sólo narrativos (*TALE-2000*). *PROLEC-SE* incluye únicamente textos expositivos, quizás debido a que la demanda de diagnóstico de dificultades de comprensión en el final de la enseñanza primaria y la totalidad de la secundaria está motivada principalmente por dificultades lectoras, especialmente en textos expositivos (Saenz y Fuchs, 2002). El texto expositivo supone un mayor reto de lectura, dado que normalmente es el utilizado para presentar ideas complejas, abstractas y relaciones lógicas (Stein y Trabasso, 1981) frente a la habitual familiaridad de los contenidos, eventos y secuencia de los textos narrativos. Características importantes ya que a partir de los últimos cursos de primaria y durante toda la secundaria se utilizan preferentemente textos expositivos, en detrimento de los narrativos (Hudson, Lignugaris-Kraft y Miller, 1993).

Dado que el Test de Procesos de Comprensión (TPC) está pensado, sobre todo, para el diagnóstico de dificultades de comprensión en situaciones educativas, decidimos también utilizar exclusivamente textos expositivos. El utilizar una mayor variedad de textos hubiera permitido una mejor estimación de la capacidad lectora de los estudiantes, similar a la realizada en los informes PISA, pero a su vez hubiera incrementado el tiempo de aplicación y el número de variables relacionadas con el producto final, por lo que su capacidad predictiva y su validez en entornos escolares quedaría ligeramente mermada.

La segunda cuestión se refiere a la decisión acerca de si se permite a los estudiantes releer el texto para responder las preguntas del test. La mayoría de los tests, con la excepción de *PROLEC* y *PROLEC-SE*, lo permiten. No permitir releer hace difícil discriminar si el error en la pregunta se debe a un problema exclusivo de recuerdo o a fallos en procesos de comprensión específicos (p. ej., formar ideas o resolver una inferencia anafórica). Dado que en la lectura inicial de los textos expositivos se suelen efectuar escasas inferencias (Britton y Gülgöz, 1991), no permitir releer los textos puede dificultar una evaluación clara de las capacidades de comprensión. Así, pensamos que permitir al lector releer el texto para contestar las preguntas tiene la ventaja de discriminar más claramente entre dificultades de recuerdo, especialmente cuando se solicita el recuerdo de infor-

maciones específicas, y dificultades en procesos de comprensión requeridas por preguntas especialmente dirigidas a evaluar alguno de esos procesos.

La tercera cuestión se refiere al formato de pregunta, es decir, respuesta abierta o de alternativa múltiple (ver una revisión del tema en Pearson y Hamm, 2005). La mayoría de los tests en español, con la excepción de PROLEC y PROLEC-SE, utilizan un formato de pregunta de alternativa múltiple. Una objeción posible es que el formato de alternativa múltiple induzca procesos de comprensión diferentes al formato de respuesta abierta, atendiendo a la distinción entre reconocimiento y recuerdo (Doolittle y Cleary, 1987; Murphy, 1982; Ward, 1980, 1982). Si bien algunos autores han encontrado que no existen, diferencias significativas en la comprensión evaluada con uno u otro formato (Bennett, Rock y Wang, 1991; Córdón y Day, 1996; van den Bergh, 1990), existen, no obstante, numerosas voces críticas hacia el uso de este tipo de preguntas, ya que permiten responder sin que el estudiante lea completamente el texto (Katz y Lautenschlager, 2001; Katz, Lautenschlager, Blackburn y Harris, 1990) o debido a que podrían activar procesos cognitivos sustancialmente diferentes (*v.g.* Berg y Smith, 1994) lo que podría generar hábitos incorrectos de lectura (*v.g.* Lubliner y Smetana, 2005). Sin embargo, en las últimas revisiones realizadas se observa que no es el formato de pregunta sino el tipo de proceso que ésta provoca, en relación con la información del texto leído, la que determina en mayor medida los procesos cognitivos implicados en su resolución (Campbell, 2005). Así, y dado que el formato de preguntas de elección múltiple resulta más fiable psicométricamente, al evitar una corrección subjetiva y más económico, en términos de aplicación y corrección, preferimos elegir este formato de pregunta, ya que esto facilita en gran medida su utilización en contextos escolares.

En resumen, el objetivo de nuestro estudio es elaborar una prueba de evaluación de la comprensión lectora bajo los siguientes supuestos: (a) basada en la evaluación de los procesos de comprensión, propuestos en los modelos teóricos actuales, (b) dirigida a escolares y orientada a la evaluación de la lectura como herramienta de aprendizaje, lo que justifica usar textos expositivos, (c) minimizar los efectos de la memoria e incitar al uso de estrategias de lectura que favorezcan la elaboración de inferencias, y (d) fácil de aplicar y corregir.

Método

Muestra

Utilizamos varias muestras de niños y niñas desde 5º de Primaria a 4º de ESO en las distintas fases de elaboración, con un reparto de las variables demográficas, tipo de centro, zona (rural *vs.* urbano) similar a la distribución de la población general. En las aplicaciones piloto utilizamos aproximadamente 750 estudiantes, divididos en siete submuestras de entre 90 y 120 estudiantes cada una.

En la aplicación definitiva la muestra estaba compuesta por 1595 estudiantes con la distribución expresada en la tabla I. Tal como se puede ver, el número de estudiantes de los diferentes cursos es aproximadamente igual, muestreando el resto de variables en una proporción aproximadamente igual al de la población general. Esta aplicación definitiva se realizó entre los meses de enero y marzo de 2005. A una parte de la muestra se le aplicó adicionalmente el test de comprensión del PROLEC-SE a fin de poder realizar un estudio de validez.

Construcción de los textos del test. Se elaboraron dos textos expositivos, “Los Pingüinos” y “Los Sioux”, cuyo contenido no tuviera solapamiento con conocimientos o temas escolares comunes. Procuramos que tanto el vocabulario (*i. e.*, frecuencia de las palabras), como la complejidad gramatical de las frases fueran adecuados para niños entre 10 y 16 años aunque los textos incluyeran palabras poco frecuentes pero

TABLA I
Distribución de la muestra de tipificación del TPC por cursos, sexo, y ámbito de aplicación

NIVEL	ÁMBITO		TIPO DE COLEGIO		SEXO		TOTAL
	Urbano	Rural	Publico	Privado	Varón	Mujer	
5º primaria (10-11 años)	133	110	126	117	111	131	243
6º primaria (11-12 años)	154	132	159	127	146	137	286
1º de ESO (12-13 años)	165	112	145	132	133	140	277
2º de ESO (13-14 años)	145	120	150	115	137	124	265
3º de ESO (14-15 años)	172	112	160	124	155	127	284
4º de ESO (15-16 años)	136	104	131	109	102	137	240

Nota: Algunos alumnos no facilitaran su nombre ni sexo lo que hace que la suma, entre las columnas referentes al sexo, sea inferior al total por curso.

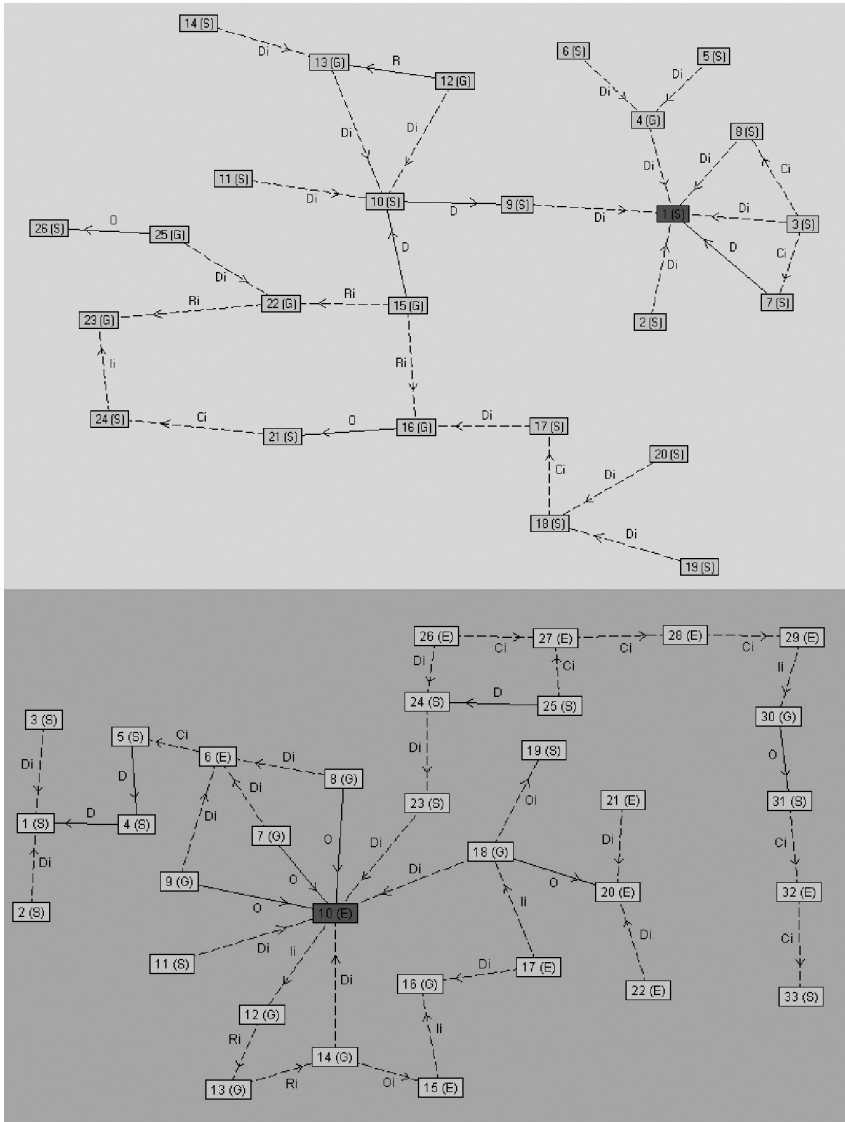
relevantes (p. ej., hábitat) cuyo significado pudiera ser inferido del contexto, al igual que ocurre en los textos académicos. Estas variables sobre la legibilidad del texto fueron también empíricamente comprobadas durante las aplicaciones iniciales, modificando aquellas palabras, frases o estructuras que hubieran resultado complejas para los estudiantes. Las variables de coherencia y organización del texto fueron estudiadas de un modo especial, dado que son los dos principales factores que afectan a la legibilidad y comprensibilidad de un texto (Beck, Omanson y McKeown, 1982; Davison y Kantor, 1982; Meyer, 2003). La organización y coherencia de ambos textos fue analizada a través de una herramienta informática de análisis de textos expositivos llamada ETAT (*Expository Text Analysis Tool*; Vidal-Abarca, Gilabert y Abad, 2002; Vidal-Abarca, Reyes, Gilabert, y Calpe, 2002). ETAT toma como unidad de análisis oraciones completas del texto, estableciéndose a posteriori relaciones semánticas existentes entre esas unidades según la clasificación de ETAT: Descripción (D), Causa (C), Razón (R), Inicia (I) y Resultado (O), pudiendo estar las relaciones explícitas o implícitas en el texto. Además, ETAT proporciona un grafo de relaciones (ver Figura 1) y una serie de estadísticos (ver Tabla II). Esta información permite ver las principales relaciones semánticas que se establecen entre las diferentes ideas del texto, si éstas son explícitas o implícitas, la estructura general y algunos índices de coherencia (p. ej. promedio y desviación estándar de relaciones por nodo). Todo ello sirve para tener una representación general de la coherencia y organización del texto y, sobre todo, ayuda en la elaboración de preguntas, al mostrar las ideas centrales y los tipos de relaciones que las unen.

Como puede verse en estos grafos ambos textos son muy similares, con relaciones semánticas variadas, si bien la que más predomina es la relación de descripción, lo cual es habitual en textos expositivos. La mayor parte de las relaciones son implícitas, lo que implica que el lector debe hacer inferencias para conectarlas. El número de unidades en Sioux, (36), resultó ser mayor que en Pingüinos (26) pese a que tenía un menor número de palabras, 469 frente a 548 de Pingüinos. Además, el promedio de relaciones por unidad también resultó mayor en Sioux; lo cual se puede apreciar comparando los grafos de ambos textos. Estos índices nos demuestran que ambos textos, aunque son similares, presentan diferentes niveles de complejidad o dificultad.

Construcción de los ítems del test. La construcción inicial de ítems se realizó a partir de los grafos. Para ilustrar el proceso de construcción utilizaremos un trozo del texto de Pingüinos.

En lo que coinciden todos los pingüinos es en el recubrimiento de su cuerpo. Tienen una espesa capa de grasa y sobre ella un abrigo de plumas cortas y muy densas colocadas de tal manera que

FIGURA 1
Grafo ETAT de los textos “Los Pingüinos” y “Los Sioux” mostrados sucesivamente



Nota: Las unidades está representadas por recuadros numerados; las flechas indican las relaciones explícitas (continuas) e implícitas (discontinuas) y el sentido de la relación.

TABLA II
Estadísticos obtenidos en ETAT de los textos LOS PINGÜINOS y LOS SIOUX

	Total U	Rel D	Rel C	Rel I	Rel R	Rel O	M rel/u	DS rel/u	% Expl
PINGÜINOS	26	18	4	1	4	2	1,12	1,24	21
SIOUX	36	16	7	4	2	7	2,10	1,00	22

Nota: Total U = total unidades; Rel = relaciones de Descripción, Causa, Inicia, Razón y Resultado (O); M rel/u = promedio de relaciones por unidad; DS rel/u = desviación estándar de relaciones por unidad; %Expl = porcentaje de relaciones explícitas sobre el total de relaciones.

forman cámaras de aire aislantes del frío ambiente. También coinciden en su solidaridad, es decir, se ayudan unos a otros, incluso sin ser de la misma familia, lo que les permite hacer frente a los paisajes hostiles y climas duros en los que viven.

Pero lo más enternecedor y original es su comportamiento reproductor. Los pingüinos Adelia, una de las especies de pingüinos antárticos, incuban sus huevos y crían a sus pequeños turnándose el macho y la hembra. Suelen formar parejas estables con bajísimo número de divorcios. Mientras uno se queda con el huevo, el otro progenitor se aleja hasta el agua para buscar comida. Después del nacimiento, ambos padres siguen compartiendo el cuidado de los polluelos y la búsqueda de comida. Comen peces y krill, un crustáceo parecido a diminutas gambas muy abundante en las aguas del Polo Sur y que sirve también de alimento para las ballenas.

A partir de este trozo de texto elaboramos cuatro tipos de preguntas de los que consta el test, y que se correspondían con los procesos comentados en la introducción: (a) formación y captación de ideas en una frase, (b) realización de inferencias-que conectan-elementos-textuales, (c) realización de inferencias-basadas en el conocimiento y (d) formación de macro-ideas.

Para elaborar preguntas sobre *captación de ideas en una frase* escogimos frases con bastantes ideas a fin de que la tarea tuviera cierta dificultad. En estas preguntas el lector debe descomponer la frase en proposiciones y analizar las relaciones semánticas y sintácticas que las proposiciones mantienen entre sí. Véase el siguiente ejemplo:

El krill es:

- a) Un pequeño pez que sirve de alimento a pingüinos y ballenas.
- b) Un crustáceo que se alimenta de peces y vive en el Polo Sur.
- c) Una especie de gamba que sirve de alimento a los pingüinos.
- d) Un crustáceo que se alimenta de gambas, al igual que los pingüinos.

Para responder adecuadamente a la pregunta el lector ha de formar estas dos ideas: *el krill se parece a las gambas* y *es un alimento para los pingüinos*, además de entender que el krill no es un pez, sino un crustáceo. Obsérvese que los distractores contienen todas ideas erróneas, así como términos de la frase, muy próximos al término en cuestión. Obsérvese igualmente que esta pregunta no se puede responder simplemente repitiendo información literal, lo que no tiene sentido cuando el lector puede consultar el texto.

La siguiente pregunta evalúa la realización de una inferencia anafórica, un caso típico *inferencia-que-conecta-elementos-textuales*.

¿Qué forma cámaras de aire en los pingüinos?

- a) Las plumas que recubren su cuerpo.
- b) La capa de grasa que recubre el cuerpo.
- c) Un abrigo de pelo que tapa su cuerpo.
- d) Una piel especial que recubre su cuerpo.

El lector debe conectar dos ideas que aparecen en el texto sobre como se cubren y aíslan los pingüinos del frío, a saber: *utilizan un recubrimiento de plumas densas y cortas*, y (las plumas) *forman cámaras de aire aislante*. Los distractores hacen referencia a ideas similares de cómo se cubren los pingüinos o elementos que podrían hacer esa función y tienen una similitud léxica con el texto.

La siguiente pregunta evalúa la capacidad para hacer una *inferencia-basada-en-el-conocimiento*.

Los pingüinos Adelia tienen un comportamiento reproductor original porque:

- a) Los pingüinos Adelia crían e incuban sus huevos.
- b) Los padres se ayudan para incubar y cuidar a las crías.
- c) Son los machos Adelia los que incuban los huevos.
- d) Cambian de pareja mientras cuidan de los huevos.

El lector deberá activar esquemas previos de conocimiento acerca del comportamiento animal, tales como que normalmente son las hembras quienes cuidan

de las crías, y no los machos, y hacer la inferencia de que el comportamiento de los pingüinos Adelia es original porque no se ajusta a este esquema. En este caso los distractores son léxicamente parecidos, manteniendo la relación entre ideas, pero sin resolver la inferencia.

Finalmente, la siguiente es una pregunta que requiere *formar una macro-idea*:

¿Qué tienen en común todos los pingüinos?

- a) El continente donde viven y el recubrimiento de su cuerpo.
- b) El recubrimiento y el tamaño de su cuerpo.
- c) La ayuda que se prestan y el tamaño de su cuerpo.
- d) El recubrimiento de su cuerpo y la ayuda que se prestan.

De las tres macro-reglas enunciadas por van Dijk y Kintsch (1983) para formar macro-ideas (*i. e.*, selección, generalización y construcción) esta pregunta requeriría aplicar la macro-regla de *selección*. El lector ha de seleccionar lo que tienen en común todos los pingüinos, omitiendo la información no esencial. Otras macro-reglas tenían que ser aplicadas en otras preguntas de macro-idea.

Para elaborar las alternativas de respuesta de los diferentes ítems nos servimos del análisis del texto y también de las respuestas dadas por 120 estudiantes en una primera aplicación piloto que constaba de 30 preguntas, 15 por texto, todas ellas con formato de pregunta abierta. En posteriores aplicaciones piloto probamos estas 30 preguntas con diferentes muestras de escolares. Para cada una de estas aplicaciones elaboramos cuatro cuadernillos distintos, cada uno con 20 preguntas, que contenían el conjunto de las preguntas. Así controlamos los efectos que el orden o la fatiga pudieran tener sobre los elementos del test. Se hicieron un total de seis aplicaciones piloto sucesivas de la prueba a diferentes subconjuntos muestrales, con una media de 110 estudiantes por aplicación. Tras cada una de estas aplicaciones se realizaron diversos ajustes de la prueba, los cuales pasamos a resumir, teniendo en cuenta diferentes índices psicométricos.

Con respecto al control de la *dificultad de las preguntas*, el análisis con ETAT y los conocimientos sobre las variables que afectan al proceso de comprensión permiten predecir en cierta medida la dificultad del ítem. Por ejemplo, una pregunta inferencial tendrá mayor dificultad si implica muchas ideas a relacionar, si éstas son implícitas, y si la distancia entre ideas es grande (Daneman y Carpenter, 1980; Singer y Ritchot, 1996). Sin embargo, esta propuesta teórica debe ser contrastada empíricamente, ya que otras variables (p. ej., una formulación ambigua o la coincidencia léxica entre el texto y la alternativa correcta) pueden alterar la dificultad de la pregunta. Durante las aplicaciones piloto controlamos el ajuste de los índices de dificultad, haciendo modificaciones en el texto, en las preguntas o en las alternativas para ajustar la dificultad a la vez que para garantizar que la respuesta se basa principalmente en la ejecución del proceso de comprensión propuesto. Así mismo, cuidamos que las alternativas incorrectas mantuvieran proporciones similares de elección.

A fin de conseguir un índice de *homogeneidad* aceptable en la prueba fuimos descartando aquellos ítems que tuvieran un bajo índice de correlación corregido con el total de la prueba. Descartamos aquellos ítems una correlación inferior a 0,30. Igualmente tuvimos en cuenta que la prueba tuviera una buena *capacidad discriminativa* entre los diferentes cursos. Para ello hicimos un análisis de contingencias, en base a las medias del ítem obtenidas por cada grupo de edad. El índice de ajuste utilizado fue el valor de la *chi-cuadrado de Pearson*, ya que en este estudio consideramos los ítems como dicotómicos, basándonos exclusivamente en el acierto o fallo al ítem. Fueron eliminados todos aquellos ítems cuyas medias no resultaron significativamente diferentes ($p < .05$).

El resultado de estos primeros estudios se concretó en una prueba de 20 ítems (10 por texto) de cuatro alternativas, representativos de los distintos procesos:

formación de ideas textuales (4), inferencias que conectan-elementos-textuales (6), inferencias basadas en el conocimiento (4) y macroideas (6).

Resultados: Análisis de las Propiedades Psicométricas del TPC

Consistencia Interna

El análisis de la consistencia interna está basado en el coeficiente *alpha de Cronbach*, obteniendo un índice global de 0,798. Cuando analizamos independientemente cada texto encontramos valores de 0,653 en Pingüinos y de 0,725 en Sioux.

Análisis de Ítems y su Fiabilidad

Los datos más destacables de los elementos que componen la escala final están plasmados en la tabla III. En esta tabla podemos observar como los índices de dificultad de los ítems ($1 - \bar{x}$) varían entre 0,13 y 0,71 estando la mayoría de ellas en torno a 0,50. Por su parte, los índices de homogeneidad nos muestran un alto nivel de consistencia interna y una aportación positiva de cada ítem al alfa de la prueba, ya que éste disminuye si eliminamos cualquiera de los ítems (ver Tabla III). Como comentamos anteriormente se tuvo especial cuidado en que los ítems diferenciaran entre los diferentes grupos de edad. Para verificar este hecho en la escala final se replicaron nuevamente los análisis basados en las tablas de contingencia por cursos e ítem. Los resultados de estas tablas muestran diferencias significativas en todos los casos. Además, las pruebas a posteriori realizadas mostraban que estas diferencias eran significativas entre cursos, siempre y cuando la distancia entre estos fuera superior a un año.

TABLA III
Estadísticos obtenidos en el análisis de ítems

Ítem	Media	<i>d.t.</i>	Homog.	α	Chi ²
Ping 1	0,53	0,499	.360	.788	100,38*
Ping 2	0,76	0,425	.381	.787	120,04*
Ping 3	0,62	0,486	.233	.794	47,41*
Ping 4	0,60	0,490	.353	.789	59,47*
Ping 5	0,50	0,500	.300	.792	44,95*
Ping 6	0,71	0,453	.319	.791	81,47*
Ping 7	0,85	0,355	.280	.792	61,38*
Ping 8	0,45	0,498	.340	.789	114,21*
Ping 9	0,87	0,342	.370	.789	58,11*
Ping 10	0,79	0,408	.389	.787	120,46*
Sioux 1	0,83	0,373	.359	.789	101,86*
Sioux 2	0,37	0,484	.286	.793	92,65*
Sioux 3	0,58	0,494	.414	.785	111,06*
Sioux 4	0,39	0,488	.380	.787	92,61*
Sioux 5	0,55	0,497	.408	.785	70,56*
Sioux 6	0,49	0,500	.333	.790	74,91*
Sioux 7	0,63	0,484	.457	.782	151,79*
Sioux 8	0,29	0,456	.362	.788	88,38*
Sioux 9	0,54	0,499	.471	.781	160,90*
Sioux 10	0,50	0,500	.439	.783	140,35*
Total	11,86	4,214			

Nota: *d.t.* = Desviación Típica; Homog= Índice de Homogeneidad; α = Alfa si eliminamos el ítem; Chi2 = Valor de Chi Cuadrado y su significación estadística con cinco grados de libertad. * $p < .001$.

Estructura Factorial

El TPC ha sido construido en base a un modelo teórico de cuatro tipos de ítems. Para descartar la posibilidad de que esos cuatro tipos de ítems dieran lugar a cuatro factores independientes realizamos un análisis factorial confirmatorio y varios exploratorios, obteniendo una escala unidimensional. Este resultado es interesante ya que los modelos cognitivos que fundamentan la prueba son modelos procesuales, no estructurales; es decir, los modelos predicen que el lector realiza procesos de forma cíclica y el producto de un ciclo es la entrada del siguiente. Con este esquema no cabe pensar en factores independientes, sino en un único factor al cual contribuyen procesos diferentes.

Validez

La *validez de contenido* fue abordada durante todo el proceso. Como dijimos, el diseño de los ítems se realizó a partir de los grafos obtenidos con ETAT y de acuerdo con el modelo cognitivo propuesto. Además, todos los ítems fueron evaluados por un conjunto de expertos, que determinaron si el proceso cognitivo implícito que realmente evaluaba el ítem se correspondía con el propuesto, con el objeto de garantizar una correspondencia con los presupuestos teóricos que sustentan la escala.

La *validez de constructo* del test se obtuvo a partir de la correlación del test con el PROLEC-SE, una escala ampliamente utilizada en el contexto español. Para ello aplicamos la prueba de comprensión PROLEC-SE a una muestra de 542 estudiantes a los que se les aplicó igualmente nuestro test. La correlación hallada entre estas dos pruebas fue 0,723 ($p < .001$), correlación especialmente alta y significativa, más si tenemos en cuenta las diferencias en el proceso de evaluación y de respuesta utilizados por cada una de las escalas.

La *validez criterial* del test fue establecida a partir de la relación que la comprensión lectora mantiene con los procesos de aprendizaje, representados en este caso por el rendimiento académico. Para ello pedimos a los centros participantes que nos proporcionaran las calificaciones de los alumnos de Secundaria en Lengua Española, Ciencias Sociales y Ciencias Naturales durante el curso anterior, transformando la nota final en cada área en una puntuación ordinal de acuerdo con la siguiente escala: Suspenso-1, Aprobado-2, Bien-3, Notable-4 y Sobresaliente-5. Tomamos únicamente las calificaciones en Secundaria porque los de Primaria únicamente reciben una puntuación dicotómica, *i. e.*, *Progresó adecuadamente vs. Necesita mejorar*. La relación entre el test y el rendimiento académico se estableció a partir del coeficiente de correlación de *Spearman*. Las correlaciones obtenidas fueron de 0,330 en *Lengua Española*, 0,390 en *Ciencias Sociales* y 0,373 en *Ciencias Naturales*, todas ellas estadísticamente significativas ($p < .01$). Hay que señalar que las correlaciones entre las puntuaciones en PROLEC-SE de aquellos alumnos a los cuales se les aplicó esta prueba y las calificaciones académicas fueron similares a las obtenidas en el caso de nuestro test, *i. e.*, 0,374, 0,487, y 0,370 para *Lengua Española*, *Ciencias Sociales* y *Ciencias Naturales*, respectivamente.

Por último, la *validez predictiva* del test se basó en el supuesto que la comprensión es un proceso evolutivo que se va adquiriendo, desarrollando y consolidando en función de la experiencia del lector. Esta afirmación implica que debe haber diferencias significativas entre cursos, acompañado de un incremento de la puntuación conforme avanza el nivel académico. Los resultados del *análisis de varianza*, muestran diferencias estadísticamente significativas en ambos textos y el total (en todos los casos, $p < .001$), así como en las pruebas a posteriori realizadas

en cada par de cursos, significativas entre casi todos los rangos de edad, excepto entre los grupos adyacentes de 2º y 3º de la ESO.

Baremación

La última parte del análisis se concretó en la elaboración de baremos por cursos y se consideró la puntuación total obtenida en el test. En la baremación del test no se observaron otras variables, tales como el sexo, tipo de centro o ámbito ya que las *pruebas t* y los *análisis de varianza* realizados al efecto no verificaron que existieran diferencias significativas en ninguna de estas variables muestrales.

Discusión y Conclusiones

Los resultados muestran cómo nuestro test se comporta de forma adecuada y cumple satisfactoriamente los objetivos marcados. Así, hemos validado una metodología de construcción de pruebas de comprensión, basada en las últimas aportaciones teóricas. Este nuevo procedimiento de construcción de pruebas añade rigurosidad, no sólo en el análisis de sus propiedades psicométricas, sino también en el análisis de los textos y en los procedimientos de creación y diseño de las preguntas. El control de la *comprensibilidad* del texto fortalece la validez del test y evita la incidencia de variables extrañas, relacionadas con la coherencia y estructura textual, que pudieran justificar las diferencias individuales más allá del proceso de comprensión. En resumen, la metodología utilizada aporta un avance en los procedimientos de generación y formulación tanto de las preguntas como de los textos, basados en los principios teóricos cognitivos actuales, lo que debería traducirse en una mejora de la validez de constructo de la prueba (Embretson y Gorin, 2001).

En cuanto a la estructura de la escala, confirmamos su unidimensionalidad, lo que no es algo excepcional, sino una constante en todas las pruebas de comprensión, ya que en ninguna de las escalas actuales se explicita la existencia de una estructura factorial. Los modelos de comprensión indican que los procesos de comprensión ocurren cíclicamente a lo largo de la lectura y se comportan necesariamente de forma interconectada; en otros términos, no es posible realizar inferencias basadas en el conocimiento si el lector no es capaz de extraer ideas, hacer inferencias textuales y, seguramente, si no forma macroideas.

Es importante resaltar que para una correcta aplicación e interpretación de los resultados del TPC es necesario ser riguroso en los procedimientos de aplicación, dado que la comprensión es un proceso estratégico en el que interactúan las variables del texto, el lector y la tarea (Kintsch y Kintsch, 2005), lo que puede producir que el mismo estímulo sea procesado de forma diferentes según la tarea (Goldman, 1997; Kintsch y Rawson, 2005; Trites y McGroarty, 2005). Esto es especialmente importante en un test como TPC en el que los lectores pueden consultar el texto para responder las preguntas. Así, una correcta aplicación de TPC implica leer los textos y luego responder a las preguntas. Los resultados podrían verse alterados si un lector pretendiera responder las preguntas sin efectuar una lectura previa (Cordón y Day, 1996; Farr, Pritchard y Smitten, 1990).

Por otra parte, el *análisis de ítems* revela una escala de alta homogeneidad, con todos los ítems altamente interrelacionados, siendo destacable su capacidad, incluso individualmente considerados, para diferenciar entre los diferentes grupos de edad. Todo ello avala la interpretación de la comprensión lectora como un proceso evolutivo en desarrollo durante la edad escolar. Dentro de este análisis de ítems también destaca el hecho de haber podido mantener moderados índices de dificultad, incluso en las preguntas más literales, pese a mantener presentes los textos durante todo el pase de la prueba. En resumen, la metodología de evalua-

ción que utiliza el TPC indica que la respuesta final a un ítem depende principalmente de que el lector complete el proceso cognitivo evaluado.

Finalmente, quisiéramos señalar que TPC es un test especialmente diseñado para su aplicación en contextos escolares. No obstante, junto a la ventaja de especialización que supone, ello conlleva una cierta limitación. No incluir textos narrativos o de otro tipo, constriñe la medida de procesos de comprensión que se obtiene. Está por ver hasta qué punto TPC correlaciona con otros tests de comprensión de espectro más amplio.

Implicaciones Educativas

El TPC es una prueba fundamentalmente desarrollada para detectar problemas de comprensión lectora en niños de 10 a 16 años de forma rápida y sencilla, lo que la convierte en una prueba básica de *screening*. La detección temprana de estos problemas puede ayudar a evitar el fracaso escolar e intervenir adecuadamente sobre los sujetos de riesgo. En los últimos informes PISA se ha puesto de relieve el gran déficit que los estudiantes españoles tienen en esta área al finalizar la secundaria (OECD, 2002; 2005), por lo que es necesario contar con procedimientos de evaluación e intervención previos. La principal aportación de esta escala frente a otras similares es tanto su fuerte fundamentación teórica, basada en las últimas aportaciones teóricas sobre comprensión, como en los altos índices psicométricos hallados.

Aunque lógicamente esta prueba no está carente de ciertas limitaciones, dada su sencillez y brevedad, está basada en una puntuación final única, sin especificar los procesos afectados, y en función de un único tipo de texto, el expositivo, dado su predominio e importancia en situaciones educativas. Posteriores trabajos en el área deberán profundizar en el análisis de los procesos de comprensión implicados, así como en las estrategias de resolución *on-line* seguidas por los estudiantes, dado que en este tipo de pruebas carecemos de información sobre los procesos de autorregulación seguidos por los estudiantes, determinantes en gran medida de los resultados que obtienen.

Referencias

- BECK, I. L., OMANSON, R. C. & MCKEOWN, M. G. (1982). An instructional redesign of reading lessons: Effects on comprehension. *Reading Research Quarterly*, 17 (4), 462-481.
- BENNETT, R. E., ROCK, D. A. & WANG, M. (1991). Equivalence of free-response and multiple-choice items. *Journal of Educational Measurement*, 28 (1), 77-92.
- BERG, C. A. & SMITH, S. (1994). Assessing students' abilities to construct and interpret line graphs: Disparities between multiple-choice and free-response instruments. *Science Education*, 78 (6), 527-554.
- BRITTON, B. K. & GÜLGÖZ, S. (1991). Using kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology*, 83 (3), 329-345.
- CAMPBELL, J. R. (2005). Single instrument, multiple measures: Considering the use of multiple item formats to assess reading comprehension. En S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 347-368). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- CORDÓN, L. A. & DAY, J. D. (1996). Strategy use on standardized reading comprehension tests. *Journal of Educational Psychology*, 88 (2), 288-295.
- COTÉ, N., GOLDMAN, S. R. & SAUL, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, 25 (1), 1-53.
- CUETOS, F., RODRÍGUEZ, B. & RUANO, E. (1996). *PROLEC (Batería de evaluación de los procesos lectores)*. Madrid: TEA.
- DANEMAN, M. & CARPENTER, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, 19 (4), 450-466.
- DAVISON, A. & KANTOR, R. N. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly*, 17 (2), 187-209.
- DE LA CRUZ, M. V. (1997). *ECL evaluación de la comprensión lectora*. Madrid: TEA Ediciones.
- DOOLITTLE, A. E. & CLEARY, T. A. (1987). Gender-based differential item performance in mathematics achievement items. *Journal of Educational Measurement*, 24 (2), 157-166.
- EMBRETSON, S. & GORIN, J. (2001). Improving construct validity with cognitive Psychology principles. *Journal of Educational Measurement*, 38 (4), 343-368.
- FARR, R., PRITCHARD, R. & SMITEN, 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.

- GOLDMAN, S. R. (1997). Learning from text: Reflections on the past and suggestions for the future. *Discourse Processes*, 23 (3) 357-398.
- GRAESSER, A. C. (en prensa). An introduction to strategic reading comprehension. En D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies*. Mahwah, NJ: Erlbaum.
- GRAESSER, A. C. & GOODMAN, S. M. (1985). Implicit knowledge, question answering, and the representation of expository text. En B. Britton & J. B. Black (Eds.), *Understanding expository text* (pp. 109-171). Hillsdale, NJ: Erlbaum.
- GRAESSER, A. C., SINGER, M. & TRABASSO, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101 (3), 371-395.
- HUDSON, P., LIGNUGARIS-KRAFT, B. & MILLER, T. (1993). Using content enhancements to improve the performance of adolescents with learning disabilities in content classes. *Learning Disabilities Research & Practice*, 8 (2), 106-126.
- KATZ, S. & LAUTENSCHLAGER, G. J. (2001). The contribution of passage and no-passage factor to item performance on the SAT reading task. *Educational Assessment*, 7 (2), 165-176.
- KATZ, S., LAUTENSCHLAGER, G. J., BLACKBURN, A. B. & HARRIS, F. H. (1990). Answering reading comprehension items without passages on the SAT. *Psychological Science*, 1 (2), 122-127.
- KINTSCH, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95 (2), 163-182.
- KINTSCH, W. (1998). *Comprehension: A paradigm for cognition*. Nueva York: Cambridge University Press.
- KINTSCH, W. & KINTSCH, E. (2005). Comprehension. En S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 71-91). Mahwah, NJ: Lawrence Erlbaum.
- KINTSCH, W. & KEENAN, J. M. (1973). Reading rate and retention as a function of the number of propositions in the base structure of the sentences. *Cognitive Psychology*, 5 (3), 257-79.
- KINTSCH, W. & RAWSON, K. A. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology*, 97 (1), 70-80.
- LÁZARO, A. J. (1999). *Prueba de comprensión lectora*. Madrid: Tea Ediciones.
- LONG, D. L., OPPY, B. J. & SEELY, M. R. (1994). Individual differences in the time course of inferential processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20 (6), 1456-1470.
- LÓPEZ, N. & ALONSO-TAPIA, J. (1996). Main idea comprehension: Training teachers and effects on students. *Journal of Research in Reading*, 19 (2), 128-153.
- LUBLINER, S. & SMETANA, L. (2005). The Effects of Comprehensive Vocabulary Instruction on Title I Students' Metacognitive Word-Learning Skills and Reading Comprehension. *Journal of Literacy Research*, 37 (2), 163-200.
- MAGLIANO, J. P., TRABASSO, T. & GRAESSER, A. C. (1999). Strategic Processing During Comprehension. *Journal of Educational Psychology*, 91 (4), 615-629.
- MEYER, B. J. F. (2003). Text coherence and readability. *Topics in Language Disorders*, 23 (3) 204-224.
- MURPHY, R. J. L. (1982). Sex differences in objective test performance. *British Journal of Educational Psychology*, 52 (2), 213.
- OAKHILL, J., CAIN, K. & YUILL, N. (1998). Individual differences in children's comprehension skill: Toward an integrated model. En C. Hulme & R. M. Joshi (Eds.), *Reading and spelling: Development and disorders* (pp. 343-367). Mahwah, NJ: Lawrence Erlbaum.
- OECD. (2002). *Conocimientos y aptitudes para la vida. Resultados de PISA 2000*. México: Santillana.
- OECD. (2005). *Informe PISA 2003: Aprender para el mundo del mañana*. Madrid: Grupo Santillana.
- PEARSON, P. D. & HAMM, D. N. (2005). The assessment of reading comprehension: A review of practices-past, present, and future. En Scott G. Paris & Steven A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 13-69). Mahwah, NJ, US: Lawrence Erlbaum Associates.
- RAMOS, J. L. & CUETOS, F. (1999). *PROLEC-SE: evaluación de los procesos lectores en alumnos de tercer ciclo de educación primaria y secundaria*. Madrid: TEA.
- RATCLIFF, R. & MCKOON, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning and Verbal Behavior*, 17 (4), 403-418.
- SAENZ, L. M. & FUCHS, L. S. (2002). Examining the reading difficulty of secondary students with learning disabilities: Expository versus narrative text. *Remedial and Special Education*, 23 (1), 31-41.
- SINGER, M. & RITCHOT, K. F. M. (1996). The role of working memory capacity and knowledge access in text inference processing. *Memory and Cognition*, 24, 733-743.
- STEIN, N. L. & TRABASSO, T. (1981). What's in a story: An approach to comprehension and instruction. En R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2, pp. 213-267). Hillsdale, NJ: Erlbaum.
- SUÁREZ, A. & MEARA, P. (1982). *CLT prueba de comprensión lectora (cloze)*. Madrid: Tea Ediciones.
- TRITES, L. & MCGROARTY, M. (2005). Reading to learn and reading to integrate: New tasks for reading comprehension tests? *Language Testing*, 22 (2), 174-210.
- URÍO, C., TORO, J. & CERVERA, M. (2000). *Escala Magallanes de lecto-escritura TALE2000*. Bilbao: Grupo Albor-Cohs.
- VAN DEN BERGH, H. (1990). On the construct validity of multiple-choice items for reading comprehension. *Applied Psychological Measurement*, 14 (1), 1-12.
- VAN DEN BROEK, P., LYNCH, J. S., NASLUND, J., IEVERS-LANDIS, C. E. & VERDUIN, K. (2003). The development of comprehension of main ideas in narratives: Evidence from the selection of titles. *Journal of Educational Psychology*, 95 (4), 707-718.
- VAN DIJK, T. A. & KINTSCH, W. (1983). *Strategies of discourse comprehension*. Nueva York: Academic Press.
- VIDAL-ABARCA, E., GILBERT, R. & ABAD, N. (2002). Una propuesta para hacer buenos textos expositivos: Hacia una tecnología del texto expositivo. *Infancia y Aprendizaje*, 25 (4), 499-514.
- VIDAL-ABARCA, E., REYES, H., GILBERT, R. & CALPE, J. (2002). ETAT: Expository text analysis tool. *Behavior Research Methods, Instruments & Computers*, 34 (1), 93-107.
- WARD, W. C. (1980). Construct validity of free-response and machine-scorable forms of a test. *Journal of Educational Measurement*, 17 (1), 11-29.
- WARD, W. C. (1982). A comparison of free-response and multiple-choice forms of verbal aptitude test. *Applied Psychological Measurement*, 6 (1) 1-11.