

Revista mexicana de investigación educativa

COMIE

comie@servidor.unam.mx

ISSN (Versión impresa): 1405-6666

MÉXICO

2004

Aldo Bazán Ramírez / Sandra Castañeda Figueiras / Silvia Macotela Flores /  
Mercedes López Valenzuela

EVALUACIÓN DEL DESMPEÑO EN LECTURA Y ESCRITURA. APORTES  
EMPÍRICOS A LA NOCIÓN DE COMPONENTES LINGÜÍSTICOS EN EL CUARTO  
GRADO DE PRIMARIA

*Revista mexicana de investigación educativa*, octubre-diciembre,  
año/vol. IX, número 023

COMIE

Distrito Federal, México

pp. 841-861

## **EVALUACIÓN DEL DESEMPEÑO EN LECTURA Y ESCRITURA**

*Aportes empíricos a la noción de componentes lingüísticos  
en el cuarto grado de primaria*

ALDO BAZÁN RAMÍREZ / SANDRA CASTAÑEDA FIGUEIRAS /

SILVIA MACOTELA FLORES / MERCEDES LÓPEZ VALENZUELA

### **Resumen:**

Se evaluaron 139 niños de Sonora, México, de cuarto grado de primaria de nueve escuelas públicas, con un instrumento de 12 series de ejercicios agrupados en tres componentes lingüísticos: reflexión sobre la lengua, lectura y escritura, derivados del plan y programas de estudios del español y los libros de texto gratuito para el cuarto grado de primaria. Los resultados muestran que el instrumento de evaluación posee validez convergente y divergente de constructo. Los resultados se discuten con base en los hallazgos obtenidos en un estudio anterior cuando los mismos niños estaban concluyendo el tercer grado de primaria (Bazán *et al.*, 2004).

### **Abstract:**

139 fourth graders from nine public elementary schools in Sonora, México, were evaluated by means of a test of twelve series of exercises grouped in three linguistic components –reflection on language, reading and writing– derived from the plan and programs of study for Spanish, and from the free text books for the fourth grade. The results show that the test has convergent and divergent validity of the construct. The discussion of the results is based on the findings from a previous study carried out when the same children were finishing the third grade (Bazán *et al.*, 2004).

---

Aldo Bazán es profesor-investigador, titular B de tiempo completo, del Departamento de Psicología del Instituto Tecnológico de Sonora. Cinco de febrero 818 sur, CP 02410, Ciudad Obregón, Sonora. CE: aldo78b@yahoo.com.mx, abazan@itson.mx

Sandra Castañeda y Silvia Macotela son profesoras titulares C de tiempo completo, de la División de Investigación y Posgrado de la Facultad de Psicología de la UNAM.

Mercedes López es profesora auxiliar del Departamento de Psicología del Instituto Tecnológico de Sonora.

**Palabras clave:** evaluación del aprendizaje, educación básica, lectura, escritura, modelos estructurales.

**Key words:** evaluation of learning, elementary education, reading, writing, structural models, Mexico.

La evaluación del aprendizaje es un tema que aún genera desacuerdos entre los especialistas y estudiosos de los problemas educativos, así como resistencias por parte de los maestros, directores de escuela y funcionarios educativos. Una de las razones de esta problemática ha sido confundir la evaluación de resultados del proceso de enseñanza-aprendizaje con el uso, casi exclusivo, de tests psicométricos o de otras pruebas estandarizadas.

Al menos, son cinco los elementos que cuestionan la evaluación del aprendizaje escolar sólo a través de la aplicación de instrumentos estandarizados, baterías y tests:

- 1) miden las ejecuciones de los niños en situaciones artificiales aplicando arbitrariamente estándares o criterios de desempeño;
- 2) omiten el uso de textos y elementos significativos para el niño y no fomentan el desarrollo de la capacidad crítica y creativa, ni el manejo de relaciones abstractas;
- 3) generan confusión de lo que es evaluación del desempeño académico con el uso y calificación de tests e instrumentos psicométricos *per se*;
- 4) reducen la práctica educativa a la elaboración, aplicación e interpretación de datos; y
- 5) sirven para determinar la promoción a grados o niveles educativos, o para determinar la aceptación o rechazo del aspirante (Díaz Barriga, 1993; Guillet y Temple, 1986; Fernández Ballesteros, 1997; Ferreiro, 1999).

Concretamente, la confusión radica en que, en lugar de evaluar el aprendizaje, lo que se hace es “medirlo” a través de la ejecución de los evaluados en tests o pruebas estandarizadas, con una marcada incidencia de la tradición psicométrica. Por ejemplo, de ninguna manera se puede afirmar que se evalúa el aprendizaje de un estudiante exclusivamente con su ejecución en una prueba estandarizada de 20 o 30 preguntas.

El malestar, entonces, no apunta al uso en sí de los instrumentos sino al que se hace de los resultados, así como a los problemas de validez ecológica

de los instrumentos que, muchas veces, incluyen actividades muy ajenas a eventos significativos y familiares a las situaciones de aprendizaje, especialmente cuando fueron elaborados para evaluar a niños de otros países o de conocimientos y habilidades de otras lenguas. Por ello, un primer aspecto que se debe delimitar es lo que implica una evaluación del aprendizaje y la medición como un elemento del proceso evaluativo.

La evaluación del aprendizaje es un proceso que implica decidir, con base en supuestos teóricos y los objetivos establecidos, el procedimiento, los instrumentos así como las condiciones para recoger información del desarrollo y logro de los objetivos del proceso de enseñanza-aprendizaje; definir el tipo de análisis de la información recogida, así como asumir juicios de valor. Es decir, se trata de un proceso de obtención de información –y de su empleo– para formar juicios que conducen a la toma de decisiones (TenBrink, 2002), que es lo que permitirá valorar los contenidos del aprendizaje (por ejemplo, programas temáticos), la forma en que los estudiantes aprenden, así como la identificación de las relaciones entre los sujetos y los contenidos que participan en el proceso educativo (Álvarez, 2001).

De acuerdo con Reimers (2003), esta evaluación es una forma particular de investigación educativa, que se ocupa de describir y explicar diversos aspectos que inciden en las oportunidades de aprendizaje, para lo cual se relacionan distintos resultados educativos con los factores que influyen en el aprendizaje de los alumnos, y estas relaciones se utilizan para proponer o establecer la efectividad de programas o políticas que podrían afectar estas condiciones.

Por otra parte, el especialista en evaluación busca recoger evidencias en relación con las características y propiedades de los fenómenos que estudia –como es el caso del aprendizaje escolar– y para ello, por lo general, utiliza una serie de instrumentos y herramientas que le permiten contar con una mejor descripción de las propiedades relevantes de su objeto de estudio (Fernandez-Ballesteros, 1997). Estas propiedades o atributos de los objetos y eventos se expresan a través de la definición de constructos que pueden ser “identificados” en la respuesta de los estudiantes en una prueba (Martínez y Moreno, 2002), o en la interacción de los alumnos con eventos u objetos en situaciones específicas, o bien, en los resultados esperados al diseñar y aplicar programas educativos. Cuando se asignan valores (puntuaciones numéricas) a esas características

o propiedades representadas en constructos o indicadores, se dice que se está midiendo un comportamiento.

Es conveniente señalar que la *medición* –obtención de valores– sólo es uno de los elementos que el especialista debe tomar en cuenta al *evaluar* –analizar– el aprendizaje de los estudiantes en un determinado periodo, en ciertos contenidos temáticos, o bien, en situaciones concretas de aprendizaje. Precisamente, la elaboración y/o uso de instrumentos para la evaluación de resultados implica que tales instrumentos recojan evidencias confiables acerca de las propiedades y atributos definidos teóricamente, de tal modo que permitan obtener valoraciones flexibles de las diferencias individuales del desempeño en actividades cercanas al mundo real de los niños, así como captar, de manera genuina, el grado de dominio en actividades de lectura y escritura, de suposiciones, inferencias y deducciones matemáticas o habilidades científicas básicas.

De igual manera, el incremento del interés en la medición de resultados de aprendizaje también demanda el énfasis en los criterios de la validez, tanto de los instrumentos como de los constructos implicados en la medición, así como del proceso mismo de la evaluación de aprendizajes.

Una forma de evaluar el aprendizaje garantizando validez ecológica es elaborar y validar instrumentos tomando en cuenta tanto los planes como los programas de estudio, las actividades, aquellas situaciones de evaluación implicadas en los libros de texto, así como los materiales de apoyo que se utilizan masivamente en el sistema educativo. Todo esto bajo una concepción global sobre la similitud entre las condiciones en las que el niño aprende, en las que ejercita estas competencias y las situaciones de evaluación; enfatizando la valoración del logro de los estudiantes mediante pruebas o situaciones basadas en el currículum (Bazán et al., 2004; Carrasco, 2003; López y Rodríguez, 2003; Martínez y Schmelkes, 1999).

Por otra parte, también cabe destacar que en las dos últimas décadas la evaluación del aprendizaje ha tenido grandes cambios, mismos que alientan su desarrollo como un área necesariamente ligada con una teoría que sustente tanto los métodos de medición utilizados como las características y propiedades a evaluar (Cole, 1993; Mislevy, 1993), tales son los casos de la evaluación adaptativa y la modelización generativa de respuestas.

En la evaluación adaptativa las pruebas no se diseñan con un carácter fijo para todas las personas sino que, más bien, los reactivos se seleccionan de un banco a partir de la información inicial, misma que se obtiene mediante el nivel de dominio de la habilidad o conocimiento que se pretende evaluar (Castañeda *et al.*, 1993). Por su parte, la modelización generativa de respuestas supone una aproximación al diseño de pruebas desde la validación de constructos, donde los ítems son generados conjuntamente con la estimación de sus parámetros psicométricos en función de los intereses de evaluación de cada ocasión (Béjar, 1993; Martínez y Moreno, 2002).

Tales constructos –descriptivos o interpretativos– se derivan de eventos y de situaciones concretas y, por lo tanto, están sujetos a criterios de validez y de significancia, puesto que éstos no son arbitrarios ni son impuestos sobre los eventos por influencia de la tradición o capricho (Kantor, 1978). Tal proceso de hacer explícita una palabra abstracta en términos de variables observables se conoce como validación del constructo. De tal manera, bajo este supuesto es posible emplear una o más medidas cuyos resultados se generalicen a una clase más amplia de indicadores que, legítimamente, empleen el mismo nombre –o constructo– (Nunnally y Bernstein, 1995).

De este modo, la validez de constructo se refiere a la garantía de que las variables empíricas son indicadores adecuados del concepto o modelo teórico a ser evaluado, sin embargo, las evidencias obtenidas al analizar la estructura de los constructos particulares sólo describen la utilidad de la teoría o enfoque con respecto a los mismos y no evalúan la verdad de una teoría en sí, ni permiten decir que los eventos o acontecimientos sobre los que se elaboran sean, en sí mismos, constructos (Kantor, 1978; Nunnally y Bernstein, 1995).

Retomando lo anterior, y con base en los planteamientos de Martínez y Moreno (2002), se puede considerar que la elaboración y/o aplicación de instrumentos para “medir” y evaluar el aprendizaje se encuentra necesariamente relacionada con la teoría que subyace a dichos instrumentos y a los constructos o variables que el especialista pretende medir, e implica tres condiciones necesarias:

- 1) delimitación teórica del constructo a medir a partir de la abstracción de alguna propiedad de los objetos o fenómenos bajo estudio;

- 2) desarrollo de un diseño de observación –o recolección– que operativice la noción teórica respecto de dichos constructos mediante la construcción de un instrumento de medición; y
- 3) aplicación de modelos matemáticos para obtener medidas de la propiedad o constructos a partir de la ejecución de determinadas tareas o respuestas recogidas con los instrumentos.

Las tres condiciones anteriores son elementales e insustituibles para aportar indicadores valiosos en la evaluación del aprendizaje. Un instrumento basado en complejos sistemas informáticos para la recolección y/o análisis estadísticos sofisticados, pero desligados de una teoría o modelos teóricos que la sustenten, poco podrá contribuir a la consolidación de una teoría de la evaluación, y a una práctica evaluativa integral que permita conocer, por lo menos, una buena parte del aprendizaje de los estudiantes.

En este estudio se analiza la validación de constructo de un instrumento de evaluación elaborado con base en el documento *Programas de estudio de español* y las actividades incluidas en los libros de texto gratuito en el cuarto grado de primaria. Cabe mencionar que en la enseñanza y aprendizaje del español en los seis grados de primaria en México, los contenidos y actividades son organizados en cuatro componentes: expresión oral, lectura, escritura y reflexión sobre la lengua (Gómez Palacio y Martínez, 2000; SEP, 2000 a y b).

Un antecedente directo es el estudio de Bazán *et al.* (2004), quienes evaluaron a 235 niños de Sonora y a 92 del Distrito Federal, con un instrumento de 17 series de ejercicios agrupados en tres componentes (reflexión sobre la lengua, lectura y escritura). Utilizando análisis factorial exploratorio y confirmatorio, los autores sólo obtuvieron un modelo de dos componentes (reflexión-lectura y escritura) y únicamente fueron confirmadas 12 de las 17 series de actividades.

Derivado del hallazgo anterior, en la presente investigación se buscó obtener validez, tanto convergente como divergente, de constructo de un instrumento para evaluar el dominio de la lengua escrita en el cuarto grado de primaria, en 12 series de ejercicios, agrupados en tres componentes lingüísticos: reflexión, lectura y escritura, con ejercicios muy similares a las 12 series utilizadas para evaluar competencias lingüísticas en el tercer grado.

Debido a la gran similitud de los contenidos que la Secretaría de Educación Pública considera para el tercer y el cuarto grados, el instrumento para ambos debe evaluar competencias similares pero en niveles de complejidad distinta. Es decir, a pesar de que analizan las mismas competencias, los dos instrumentos deben variar en el nivel de complejidad que se demanda de los niños en cada grado escolar. De tal manera, para lograr esta diferencia cualitativa, se consideraron dos aspectos fundamentales:

- 1) incluir más elementos en la tarea o ejercicio que se expone al niño, es decir, un criterio más cuantitativo o referido a aspectos morfológicos, y/o
- 2) incluir en la tabla de categorías criterios de logro más elevados para la serie en cuestión, es decir, criterios que enfatizan el aspecto cualitativo o el nivel de complejidad funcional en que se espera se ejecute la tarea.

## **Método**

### **Participantes**

De los 167 niños provenientes de nueve grupos escolares de Ciudad Obregón, Sonora, quienes fueron evaluados en el estudio de Bazán *et al.* (2004), en esta investigación sólo fueron incluidos 139 niños, los cuales llegaron hasta el último bimestre del cuarto grado de primaria, tuvieron al menos 60% de asistencia en el año escolar 2002-2003 y cuyos padres completaron la información sobre apoyo e interés familiar solicitados para otro estudio realizado con estos niños y sus familias.

El 48% de los participantes fueron varones (66) y el porcentaje restante mujeres (73). En cuanto al promedio de edad, 27% (37) de los participantes tenía un promedio de nueve años; 68% (95) uno de 10 años y 5% (7) de 11 años de edad.

### **Instrumento**

El proceso de construcción del instrumento consistió en delimitar las competencias que los niños debían mostrar, los indicadores de ejecución y los criterios de evaluación para cada competencia, en tres componentes: reflexión sobre la lengua, lectura y escritura así como sus subcomponentes (reflexión sobre los códigos de comunicación oral y escrita, comprensión lectora, producción de textos, funciones de la escritura).

Con apego a los programas de estudio de español y los libros tanto para el maestro como el de texto considerados para el cuarto grado de primaria, se elaboró una variedad de ejercicios para cada una de las 12 competencias formando, de esta manera, 12 series de ejercicios o actividades. Esta primera versión fue presentada a dos especialistas en enseñanza de español, lo que sirvió para modificar algunas actividades y complementar otras. Cabe mencionar que en el libro de texto de cuarto grado (SEP, 2000b) no se incluyen actividades para la competencia “escribir sonidos de animales a partir de un ejemplo”, en contraste con el libro de texto de tercero. De esta forma, para que la evaluación en cuarto pudiese incluir las mismas 12 competencias resultantes cuando los niños fueron evaluados en tercero, se construyeron ejercicios en esta competencia, parecidos a los utilizados en ese grado, aunque con mayor nivel de complejidad.

Posteriormente, el instrumento fue aplicado a una muestra piloto conformada por 156 niños del segundo bimestre de quinto grado. Los evaluadores tomaron nota de los problemas o dudas más frecuentes de los niños con el propósito de mejorar las instrucciones o aquellas actividades donde los alumnos tenían problemas, no entendían o se confundían. Con excepción de la quinta serie –conformada por actividades de responder cuatro preguntas a partir de un texto informativo– en todas las series se obtuvieron índices de consistencia interna superior a 0.69. Si embargo, no se encontró validez de constructo al analizar estos datos con un programa de ecuaciones estructurales. Un análisis profundo de estos hallazgos condujo a concluir que una de las razones principales se centra en la dificultad de discriminación o divergencia de las 12 series de actividades con respecto a los componentes o factores.

Con base en las observaciones de las especialistas, los resultados y observaciones en la prueba piloto, se rediseñaron el instrumento, la tabla de categorías y la hoja de registro. Una de las ventajas de este instrumento es que cada serie de ejercicios, por cada categoría o competencia, incluye al menos dos actividades, mientras que ninguna serie contiene más de cinco.

### Procedimiento

En los meses de mayo y junio de 2003, dos evaluadores por cada grupo escolar aplicaron, de manera colectiva, el instrumento a los 139 niños, en dos sesiones de una hora de duración cada una. Se aseguraron de que

todos los alumnos comprendieran las instrucciones escritas antes de dar inicio a los ejercicios del instrumento y debían proporcionar apoyo a los estudiantes cuando se identificaba que tenían algún problema o duda.

### Categorización

Se consideró como categoría a cada una de las competencias que, de acuerdo con los planes y programas de la SEP, se pretende lograr en cada componente o bloque. Cada una incluyó al menos dos series de actividades, criterios específicos de logro para definir los requisitos mínimos de la ejecución que responde a la demanda de la tarea (competencia) y la valoración de las respuestas en términos de asignación de puntajes. La escala de calificación de la ejecución empleada fue desde 0 –actividad no realizada, la respuesta no presenta ninguno de los criterios de logro especificados– hasta 4 –realización de la actividad cumpliendo todos los criterios de ejecución.

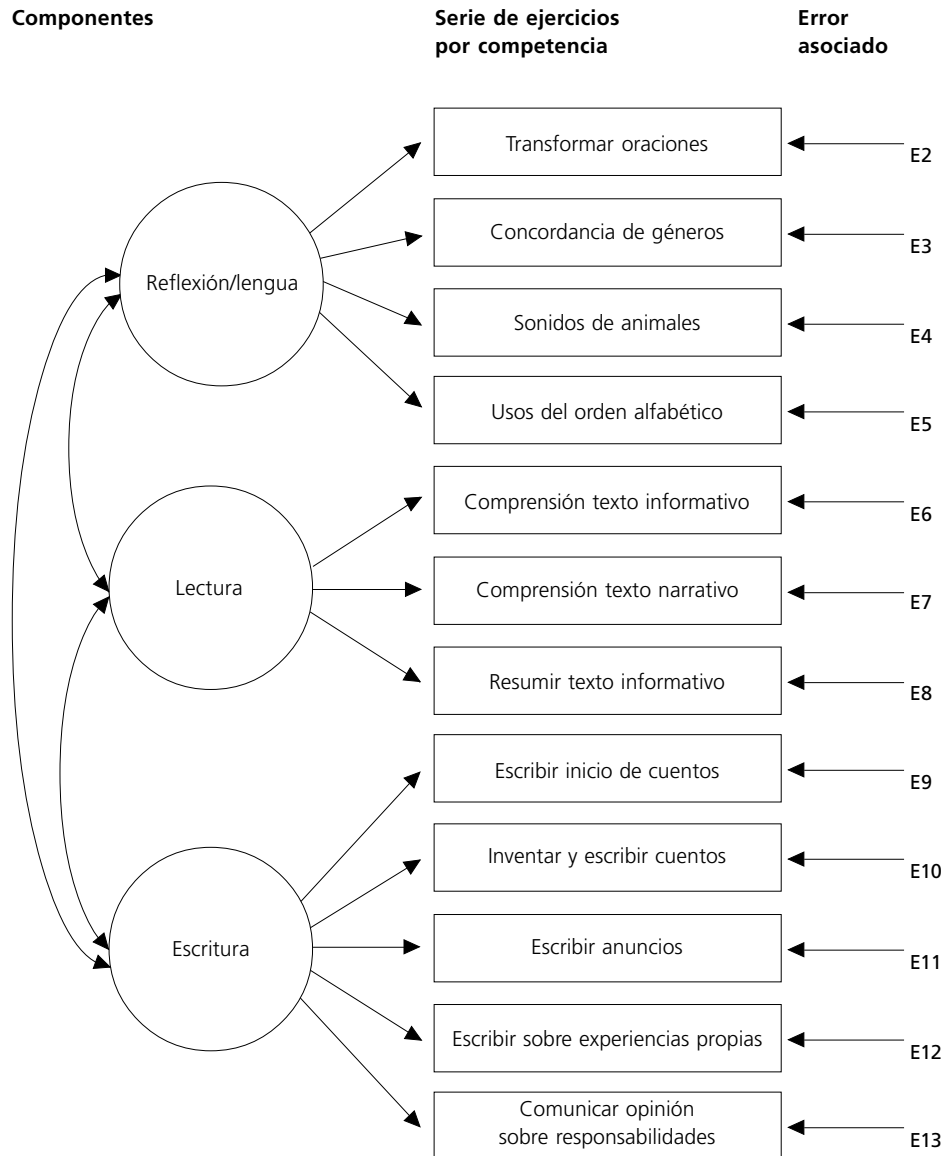
### Análisis de datos

Para obtener indicadores de consistencia interna del instrumento se obtuvieron índices alpha de Cronbach por cada una de las 12 series de actividades; posteriormente, se obtuvo el índice general de consistencia interna del instrumento tomado con base en los promedios obtenidos en cada serie después de eliminar los índices inferiores a 0.40; es decir, se consideraron los ejercicios con mejores índices de consistencia interna. En el caso de series con dos, se tomaron ambos ejercicios para obtener los promedios o índices por cada serie.

Estos promedios de cada serie fueron ingresados como indicadores en la base de datos del programa EQS6 (Bentler y Wu, 2002), con la finalidad de obtener indicadores de validez, tanto divergente como convergente, de la evaluación del cuarto grado de primaria con 12 series agrupadas en tres componentes.

Se propuso un modelo hipotético acerca de cómo se organizarían las ejecuciones de los niños en tareas de reflexión sobre la lengua-lectura y escritura (ver figura 1) y se buscó probarlo a través del análisis factorial confirmatorio. Precisamente, una estrategia útil para poner a prueba modelos hipotetizados –comparándolo con datos empíricos y analizados sistemáticamente– es el análisis factorial confirmatorio (AFC) (Byrne, 1994; Loehlin, 1992), mismo que permite determinar, de manera simultánea, validez convergente y divergente de constructos.

FIGURA 1  
*Modelo hipotético de análisis factorial confirmatorio  
 de 12 series de ejercicios en tres componentes*



## Resultados

En lo que se refiere a la consistencia interna del instrumento, en la tabla 1 se describe el índice de alpha de Cronbach obtenido en cada serie de actividades. La quinta serie –conformada por actividades de responder cuatro preguntas a partir de un texto informativo– fue la que obtuvo el coeficiente más bajo (0.53) en relación con las otras 11, mismas que obtuvieron índices de consistencia superior a 0.62.

De acuerdo con esta tabla, si se elimina la actividad 2 (RORAC2) de la primera serie, su consistencia general mejora notablemente (0.63) y, si se elimina la primera actividad de la tercera serie (RSONI1), su consistencia general también se incrementa (0.68). Por ello, en análisis posteriores se eliminaron estos dos ejercicios. El índice general de consistencia interna del instrumento tomando como indicadores el promedio obtenido en cada una de las 12 series fue 0.73.

Por otra parte, no fue confirmado el modelo que supone que las 12 series de ejercicios se organizan en tres componentes lingüísticos; en cambio, sólo se obtuvo uno de dos: reflexión y lectura-escritura, que se muestra en la figura 2.

El modelo resultante tiene bondad de ajuste ( $\chi^2$  cuadrada = 69.46,  $P = 0.08$ , indicador práctico de bondad de ajuste CFI=0.90) y permite identificar la configuración de dos grupos de factores relacionados con los componentes lingüísticos y sus respectivos indicadores (series de ejercicios o variables). Con excepción de la serie reflexión sonido de animales (RSONI), todas recibieron pesos factoriales significativos de su respectivo factor o componente lingüístico.

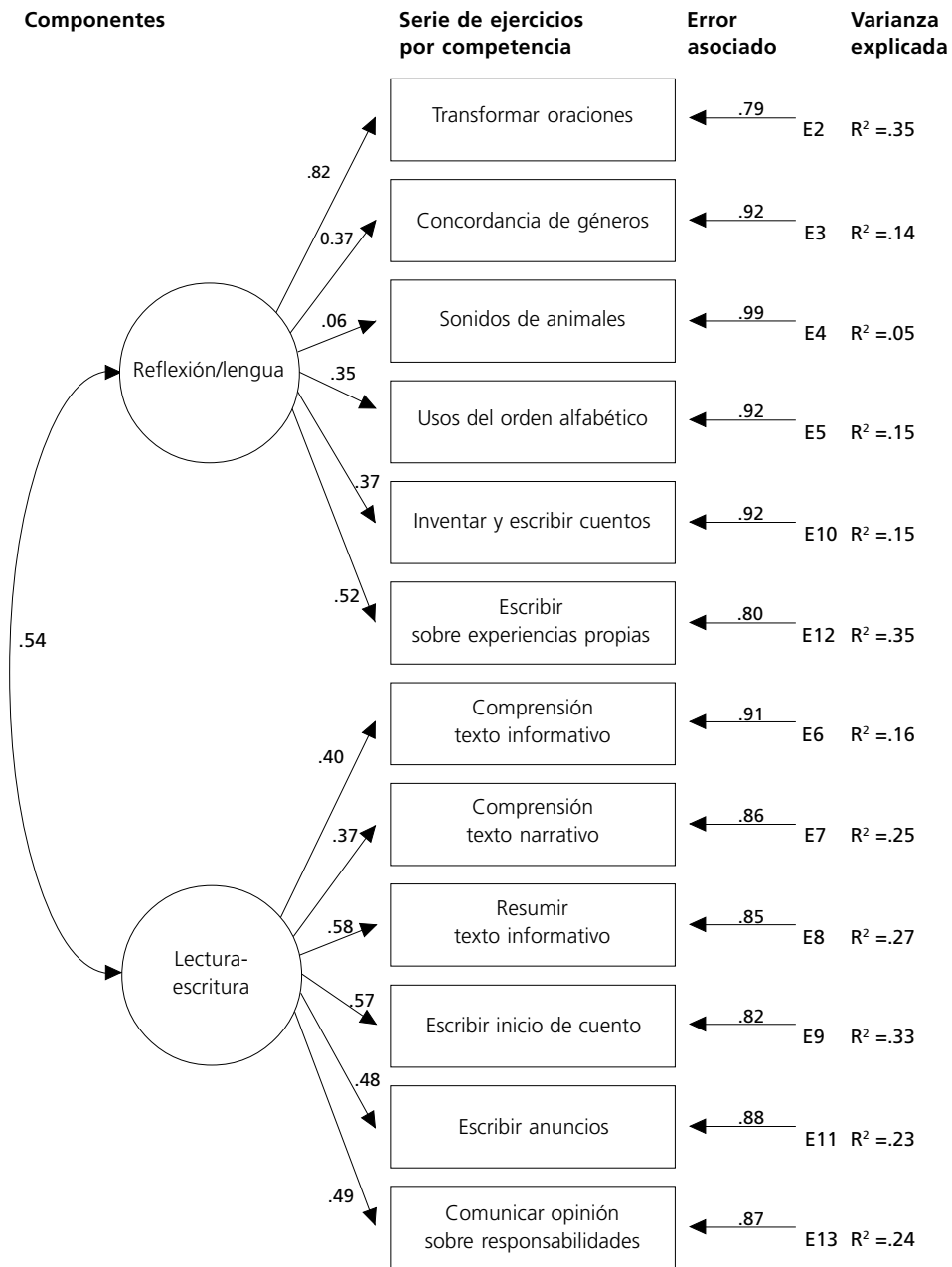
Lo anterior denota que la evaluación de la lengua escrita en el cuarto grado de primaria, en dos componentes lingüísticos, tiene validez convergente de constructo. Asimismo, la covarianza entre los dos componentes (constructos o variables latentes) es moderadamente significativa pero inferior a los pesos factoriales entre los componentes y sus indicadores (componentes y series de ejercicios), lo anterior quiere decir, además, que el instrumento permite evaluar en dos componentes lingüísticos diferentes.

Al igual que cuando estos estudiantes fueron evaluados al final del tercer grado, en cuarto se obtuvo un instrumento de dos componentes, con buena validez convergente y divergente de constructo, pero dos variables (lectura de textos informativos y lectura de textos narrativos) que en tercero formaban parte del componente reflexión sobre la lengua-lectura; en cuarto grado se agruparon junto con indicadores del componente escritura.

**TABLA 1**  
*Índice de consistencia interna en cada serie de actividades*

Serie de ejercicios (competencias)	Si se excluye la tarea	Coeficiente alpha	
		Por cada tarea	Para cada serie
Transformación de oraciones y apreciación del cambio en el sentido	RORAC1	0.5193	0.58
	RORAC2	0.6308	
	RORAC3	0.4852	
	RORAC4	0.5009	
	RORAC5	0.5042	
Uso de la concordancia de género, número, persona y tiempo en las oraciones	RGENE1	0.6732	0.76
	RGENE2	0.7009	
	RGENE3	0.6922	
	RGENE4	0.7344	
Interpretación y uso de palabras y frases a partir del significado sintáctico-semántico	RSONI1	0.6842	0.67
	RSONI2	0.5736	
	RSONI3	0.6035	
	RSONI4	0.6356	
	RSONI5	0.5734	
Conocimiento del alfabeto y uso del orden alfabético	RORDEN1	0.7689	0.83
	RORDEN2	0.8219	
	RORDEN3	0.7482	
	RORDEN4	0.7648	
Comprensión de texto informativo	TEXINF1	0.4887	0.53
	TEXINF2	0.4429	
	TEXINF3	0.4174	
	TEXINF4	0.4868	
Comprensión de un texto narrativo	TEXNAR1	0.6661	0.75
	TEXNAR2	0.7294	
	TEXNAR3	0.7069	
	TEXNAR4	0.6639	
Escribir un resumen a partir de la lectura de un texto informativo	ERESU1	.	0.77
	ERESU2	.	
Composición de cuentos (iniciar)	EINICUE1	0.8459	0.87
	EINICUE2	0.8071	
	EINICUE3	0.8049	
Composición de cuentos (escribir)	EINVCUE1	.	0.62
	EINVCUE2	.	
Elaboración de textos informativos	EANUN1	.	0.86
	EANUN2	.	
Relato escrito de experiencias	EEXPE1	.	0.80
	EEXPE2	.	
Escritura para comunicar sentimientos, opinión y justificación	EOBLI1	0.8559	0.89
	EOBLI2	0.8378	
	EOBLI3	0.8517	

FIGURA 2  
*Modelo resultante de análisis factorial confirmatorio de la organización de 12 series de ejercicios en dos componentes*



De igual manera, dos variables (inventar cuentos a partir de dibujos de personajes, y escribir de experiencias propias) que en tercer grado formaban parte del factor escritura, en cuarto se agruparon junto con cuatro indicadores de reflexión sobre la lengua.

De esta forma, se obtuvo que el desempeño de los niños de cuarto grado en pruebas de ejecución se estructura en dos factores o componentes: reflexión sobre la lengua y lectura-escritura (este último conformado con tres series de actividades de lectura y tres de escritura, inicialmente consideradas). Asimismo, inventar cuentos y escribir sobre experiencias propias, que en teoría forman parte del componente escritura y que en tercer grado sí estaban en él, en el cuarto se ubican en el de reflexión sobre la lengua.

Por otra parte, análisis comparativos mostraron que en el componente reflexión sobre la lengua no existen diferencias significativas por sexo, pero en el conformado por ejercicios de lectura y escritura sí las hay en favor de las niñas, quienes obtuvieron una media de 1.98 en comparación con los varones, con una media de 1.77.

Un análisis no paramétrico con la prueba Kruskal Wallis obtuvo una chi cuadrada ( $\chi^2$ ) = 25.34 con  $P = .001$  para la comparación de rangos por escuelas en reflexión sobre la lengua, y una  $\chi^2 = 24.21$  con  $P = .002$  para la comparación de rangos por escuela en lectura y escritura. Es decir, que existen diferencias significativas en los promedios obtenidos en reflexión sobre la lengua, como en lectura-escritura, dependiendo de la escuela de procedencia de los niños. Asimismo, un análisis de varianza de una vía mostró que existen diferencias significativas entre grupos, tanto en la evaluación en reflexión sobre la lengua ( $F = 3.43$ ,  $P = .001$ ), como en la correspondiente a lectura y escritura ( $F = 2.83$ ,  $P = .006$ ).

En la tabla 2 se presentan los promedios obtenidos por los niños en ambos componentes (reflexión, y lectura y escritura) y la desviación estándar, por cada una de las nueve escuelas de procedencia.

### Discusión

Un primer aspecto a considerar es la posibilidad de obtener un instrumento que proporcione información sistemática complementaria acerca del desempeño de los estudiantes en tareas que demanden que desarrolle y muestre las competencias –saber hacer– de acuerdo con lo que se espera que hagan los niños en situaciones similares a las del aprendizaje y ejercicio de dominios en la materia de español.

TABLA 2

*Promedios obtenidos en reflexión sobre la lengua  
y en lectura y escritura, desviación estándar y número de sujetos por escuela*

Escuela		Reflexión sobre la lengua	Lectura y escritura
1	Media	1.5095	2.0814
	N	22	22
	Desviación estándar	.51738	.69151
2	Media	2.2340	1.8761
	N	13	13
	Desviación estándar	.56712	.39826
3	Media	2.0271	2.1271
	N	20	20
	Desviación estándar	.56830	.68615
4	Media	2.0817	1.8397
	N	26	26
	Desviación estándar	.58446	.46090
5	Media	1.6218	2.0855
	N	13	13
	Desviación estándar	.51943	.63343
6	Media	1.7326	1.8484
	N	12	12
	Desviación estándar	.57802	.34290
7	Media	2.0625	1.2616
	N	6	6
	Desviación estándar	.38886	.25736
8	Media	1.7143	1.6012
	N	14	14
	Desviación estándar	.47051	.39922
9	Media	1.7788	1.6987
	N	13	13
	Desviación estándar	.47641	.47593
<b>Total</b>	Media	1.8582	1.8843
	<b>N</b>	<b>139</b>	<b>139</b>
	Desviación estándar	.57323	.56437

Lo anterior implica que los elaboradores de los instrumentos de evaluación de aprendizajes cercanos al currículum y a las experiencias significativas de aprendizaje diseñen actividades cualitativamente iguales a las de los programas y libros de texto con las que se supone que los niños trabajan en el aula –que cumplan los mismos objetivos, demanden las mismas competencias, y que traten del mismo tema– pero morfológicamente diferentes, es decir que no se incluyan exactamente las mismas actividades de los libros.

Se busca que los instrumentos de evaluación elaborados así recojan elementos que forman parte de la situación de aprendizaje y que permitan a los niños resolver actividades novedosas, de manera variada y efectiva, evitando utilizar situaciones artificiales o descontextualizadas para ellos. Este tipo de instrumentos implica, muchas veces, que sus elaboradores se despojen de su propio enfoque teórico particular y diseñen actividades apegadas a los postulados teóricos que subyacen a los programas de estudio y los materiales educativos derivados de tales propuestas. De tal manera que, sin asumir una posición militante con respecto al marco teórico de tales propuestas y materiales educativos, los evaluadores puedan comprender y delimitar los constructos incluidos.

Esta forma de proceder no implica, de ninguna manera, una postura ecléctica, más bien, se aproxima a la modelización generativa de respuestas en la elaboración de instrumentos de evaluación desde la validación de constructos, como mencionamos en la primera sección de este artículo (Béjar, 1993).

Esto es importante en el intento de proponer instrumentos de evaluación con actividades muy similares a las incluidas en los programas de estudio y los materiales educativos, para suponer que se está evaluando, de manera confiable y válida, al menos una parte del aprendizaje de los niños en la materia de español –la otra parte de la información podrá encontrarse, entre otros, en los legajos y carpetas personales, las observaciones de la maestra o las valoraciones. En este contexto, los resultados de este estudio muestran que las 12 series de ejercicios se agrupan sólo en dos componentes o factores y que éstos divergen significativamente, es decir, se obtuvo un modelo satisfactorio de validación de constructo, lo cual sugiere que se cuenta con un instrumento de evaluación de dos componentes lingüísticos diferentes en el cuarto grado: reflexión sobre la lengua y lectura-escritura.

Sin embargo, el modelo resultante sólo explicó 5% de la variabilidad de la serie “sonidos de animales”; que forma parte del subcomponente reflexión sobre los códigos de comunicación oral y escrita y demanda que el niño escriba sobre los sonidos de animales a partir de un ejemplo, incluyendo una correspondencia entre los significado global y local (SEP, 2000a). Si bien es cierto que esto no afecta la validez convergente del constructo o componente reflexión sobre la lengua en el modelo resultante, este mismo sugiere que la variable “sonidos de animales” no es, al parecer, un buen indicador de tal constructo. Una explicación alternativa podría ser que el desempeño de los niños en esa serie de ejercicios es inferior en comparación con el desempeño en las otras 11 debido, probablemente, a la ausencia de este tipo de actividades en situaciones de aprendizaje.

El segundo aspecto a tomar en cuenta en esta sección se relaciona concretamente con la estructuración de competencias lingüísticas en componentes, siguiendo las propuestas del programa para la enseñanza del español vigente. Los datos sugieren que no se confirma el supuesto de que la ejecución de los niños en evaluaciones de desempeño se estructuran de manera similar en los mismos componentes, en tercero y cuarto grados de la escuela primaria, tal como propone la SEP. Al igual que en el tercer grado, en cuarto grado también la ejecución de los niños en las 12 series de ejercicios, sólo se agruparon factorialmente en dos componentes. Sin embargo, en cuarto, la organización de las 12 series fue distinta a las mismas en el tercer grado.

En un estudio anterior, Bazán y López (2002), al evaluar a 138 niños que habían culminado el tercer grado de primaria, con 36 series de ejercicios no encontraron validez divergente de constructo entre cuatro componentes lingüísticos: hablar y escuchar, lectura, escritura y reflexión sobre la lengua, pero encontraron un modelo alternativo que sugiere reorganizar la ejecución de los niños en cuatro nuevos factores o componentes. Los hallazgos en el presente estudio y los obtenidos por Bazán *et al.*, (2004) y por Bazán y López (2002) permiten suponer, al menos, dos problemas fundamentales en la organización de la enseñanza y evaluación de la lengua escrita, en componentes lingüísticos:

- 1) Al menos en tercer y cuarto grados de primaria, las actividades elaboradas para la evaluación de aprendizajes de la lengua escrita no se organizan siguiendo los bloques o componentes propuestos, ni cuando

se toman los cuatro originales, ni cuando se toman sólo tres, una vez excluido el bloque expresión oral; estos hallazgos sugieren que existen deficiencias de validez de constructo, puesto que, al parecer, da lo mismo organizar las actividades de enseñanza-aprendizaje del español, en dos o más componentes.

- 2) La diferencia de estructuración de indicadores en tercero y cuarto grados podría estar indicando cambios cualitativos en la adquisición de las competencias asociadas con la lectura y escritura, derivados del incremento en complejidades –cualitativas y morfológicas– de la enseñanza de la lengua escrita, aunque el programa de la SEP considere que los contenidos deban ser iguales para ambos grados escolares. Esto quiere decir que, probablemente, el desempeño de los niños en contextos tanto de enseñanza como de evaluación de resultados de aprendizaje no se organizan de manera estrictamente igual en todo los niños, ni en todos los grados escolares de primaria.

Coincidentemente, algunos de los problemas de proximidad y diferenciación de las actividades incluidas en componentes en la propuesta para la enseñanza del español, a partir de la reforma curricular de 1993 en México, han sido señalados por Carrasco (2000, referido por López y Rodríguez, 2003), y por López y Rodríguez (2003). Sin embargo, y como se ha mencionado también en la primera sección de este trabajo, las evidencias obtenidas al analizar la estructura de los componentes lingüísticos (constructos) sólo describen algunos aspectos en la pertinencia del enfoque teórico que supone la estructuración en componentes, del aprendizaje –y la evaluación– del español en la escuela primaria.

En este sentido, el aporte del presente trabajo fue mostrar que es posible evaluar el aprendizaje con una perspectiva ligada con una necesaria reflexión teórica sobre el quehacer evaluativo, pero con una posible y saludable independencia de los enfoques teóricos que dan soporte a la planeación y ejecución de contenidos de aprendizaje. El otro aporte radica en que es posible evaluar resultados de aprendizaje en español, con actividades muy familiares a las incluidas en los programas de estudio y en los materiales educativos y que, a la vez, la evaluación pueda proporcionar soporte empírico a los constructos incluidos en dichos programas de estudio. No sólo se cuenta con un instrumento que cumple con criterios de confiabilidad y validez –y que permite identificar cuáles competencias y/o

habilidades dominan los niños y en qué nivel y qué componentes o contenidos presentan, en general, dificultades en relación con el nivel de ejecución de los niños— sino que, además, invita a reflexionar sobre los constructos que subyacen en la enseñanza del español así como en la organización de los contenidos de acuerdo con criterios morfológicos o taxonómicos cualitativamente jerarquizados, como es el caso de la organización de contenidos por componentes.

Cabe mencionar que esta propuesta sólo abarca la evaluación de ejecuciones de aprendizaje con indicadores incluidos expresamente en los programas y materiales educativos sugeridos para la enseñanza del español en cuarto grado, pero es menester considerar que en la evaluación del aprendizaje deberá considerarse, también, la influencia de distintas variables en el proceso educativo (Bazán *et al.*, en dictamen; Bazán, Osuna y Ross, 2003; Buckner, Bassuk y Weinreb, 2001; Cervini, 2003; Leseman, y de Jong, 1998; LLECE-UNESCO, 2000; Mella y Ortiz, 1999; Ravela, 2003; Resnick y Harwell, 2000; Schmelkes, 1997).

### Agradecimiento

Los autores agradecen al licenciado Pablo Reyes Castro, por su apoyo en la corrección de las figuras incluidas en este artículo. Esta investigación se realizó como parte de los estudios de doctorado en Psicología del primer autor.

### Referencias bibliográficas

- Álvarez, J. M. (2001). *Evaluar para conocer, examinar para excluir*, Madrid: Morata.
- Bazán, A. y López, M. (2002). “Los componentes en la adquisición de la lengua en la escuela primaria mexicana: un caso de validación de constructos”, *Revista Mexicana de Investigación Educativa*, vol. 7, núm. 16, pp. 603-625.
- Bazán, A.; Osuna, B. y Ross, G. (2003). “Interés familiar, opinión del maestro, y desempeño en lectura y escritura en niños de primaria”, *Revista Mexicana de Psicología*, vol. 20, núm. 2, pp. 255-264.
- Bazán, A. *et al.* (2004). “Aportes para la evaluación de tres componentes lingüísticos en la enseñanza de la lengua escrita en el tercer grado de primaria”, *Revista Latina de Pensamiento y Lenguaje*, vol. 12, núm. 1, pp. 37-67.
- Bazán, A. *et al.* (en dictamen). “Structural modelling of endogenous and exogenous variables in the acquisition of written language in mexican children”, *Early Years*, Inglaterra.
- Béjar, I. (1993). “A generative approach to psychological and educational measurement”, en N. Frederiksen, R. Mislevy e I. Bejar (eds.), *Test theory for a new generation of test*, New Jersey: Lawrence Erlbaum, pp. 323-357.

- Bentler, P. y Wu, E. (2002). *EQS 6 for Windows. User's Guide*. Encino: Multivariate Software Inc.
- Buckner, J.; Bassuk, E. y Weinreb, L. (2001). "Predictors of academic achievement among homeless and low-income housed children", *Journal of School Psychology*, vol. 39, núm. 1, pp. 45-69.
- Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: basic concepts, applications, and programming*, California: SAGE Publications.
- Carrasco, A. (2003). "Experiencia exitosa de enseñanza de la lectura en la escuela primaria", en *Memoria electrónica del VII Congreso Nacional de Investigación Educativa*. Guadalajara: COMIE.
- Castañeda, S. et al. (1993). "Un marco de trabajo neurocomputacional para el estudio de la estructuración del conocimiento a partir de lo leído", *Revista Latina de Pensamiento y Lenguaje*, vol. 1, núm. 2, pp. 201-232.
- Cervini, R. (2003). "Relación entre composición escolar, proceso escolar y el logro en matemática del nivel secundario de Argentina", *Revista Electrónica de Investigación Educativa*, vol. 5, núm. 1 [consultado el 20 de abril de 2004 en <http://redie.ens.uabc.mx/vol5no1/contenido-cervini2.html>].
- Cole, N. (1993). "Comments on chapters 1-3", en N. Frederiksen, R. Mislevy y Béjar, I. (eds.), *Test theory for a new generation of test*, Nueva Jersey: Lawrence Erlbaum, pp. 72-77.
- Díaz Barriga, A. (1993). "El problema de la teoría de la evaluación y cuantificación del aprendizaje", en A. Díaz Barriga (comp.) *El examen: textos para su historia y debate*, México: UNAM-CESU, pp. 304-314.
- Fernández Ballesteros, R. (1997). "Evaluación psicológica y tests", en A. Cordero (coord.) *La evaluación psicológica en el año 2000*, Madrid: TEA, pp. 1-26.
- Ferreiro, E. (1999). *Cultura escrita y educación. Conversaciones con Emilia Ferreiro*, México: Fondo de Cultura Económica.
- Gillet, J. W. y Temple, Ch. (1986). *Understanding reading problems. Assessment and instruction*, Boston: Little Brown and Company.
- Gómez Palacio, M. y Martínez, A. (coord.) (2000). *La adquisición de la lectura y la escritura en la escuela primaria. Lecturas*, México: Secretaría de Educación Pública-Programa Nacional de Actualización Permanente.
- Kantor, J. R. [1959] (1978). *Psicología interconductual. Un ejemplo de construcción científica sistemática*. México: Trillas.
- Leseman, P. y de Jong, P. (1998). "Home literacy: opportunity, instruction, cooperation and social-emotional quality predicting early reading achievement", *Reading Research Quarterly*, vol. 33, núm. 3, pp. 294-318.
- LLECE-UNESCO (2000). *Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados, para alumnos de tercer y cuarto grados de la educación básica. Segundo Informe*, Santiago de Chile: UNESCO.
- Loehlin, J. (1992). *Latent variable models: an introduction to factor, path, and structural analysis*, New Jersey: Lawrence Erlbaum.

- López, G. y Rodríguez, M. (2003). "La evaluación alternativa: oportunidades y desafíos para evaluar la lectura", *Revista Mexicana de Investigación Educativa*, vol. 8, núm. 17, pp. 67-98.
- Martínez, F. y Schmelkes, S. (1999). "Aseguramiento de la calidad de las pruebas de estándares nacionales para la educación primaria, de la Secretaría de Educación Pública", en *Memoria electrónica del V Congreso Nacional de Investigación Educativa*. Aguascalientes: COMIE.
- Martínez, R. y Moreno, R. (2002). "Integración de teoría sustantiva, diseños de pruebas y modelos de análisis en la medición psicológica", en A. Bazán y A. Arce (eds.), *Estrategias de evaluación y medición del comportamiento en Psicología*, Ciudad Obregón: ITSON-UADY, pp. 87-119.
- Mella, O. y Ortiz, I. (1999). "Rendimiento escolar. Influencias diferenciales de factores externos e internos", *Revista Latinoamericana de Estudios Educativos*, vol. 29, núm. 1, pp. 69-92.
- Mislevy, R. (1993). "Foundations of a new test theory", en N. Frederiksen, R. Mislevy y Béjar, I. (eds.), *Test theory for a new generation of test*, Nueva Jersey: Lawrence Erlbaum, pp. 19-39.
- Nunnally, J. y Bernstein, I. (1995). *Teoría psicométrica*, México: McGrawHill.
- Ravela, P. (2003). "Procesos escolares, contextos socioculturales y logros en el aprendizaje. Una visión integradora para la investigación educativa", en COMIE, *VI Congreso Nacional de Investigación Educativa. Conferencias Magistrales*, México: COMIE, pp. 97-128.
- Reimers, F. (2003). "El contexto social de la evaluación educativa en América Latina", *Revista Latinoamericana de Estudios Educativos*, vol. 33, núm. 3, pp. 9-52.
- Resnick, L. B. y Harwell, M. (2000). "Instructional variation and student achievement in standards-based education district", en *CS Technical Report 522*, Los Angeles: Center for the Study of Evaluation.
- Schmelkes, S. (1997). *La calidad en la educación primaria. Un estudio de caso*, col. Biblioteca del Normalista, México: SEP.
- Secretaría de Educación Pública (2000a). *Programas de estudio de español. Educación primaria*, México: SEP.
- Secretaría de Educación Pública (2000b). *Español. Cuarto grado. Actividades*. México: SEP.
- TenBrink, T. (2002/1986). "Evaluación", en J. Cooper (editor), *Estrategias de enseñanza. Guía para una mejor instrucción*, México: Limusa, pp. 499-558.

**Artículo recibido:** 4 de mayo de 2004

**Aceptado:** 12 de octubre de 2004