Revista Interamericana de Psicología/Interamerican Journal of Psychology - 2009, Vol. 43, Num. 3 pp. 518-531

Evaluación de la Escritura Mediante Rúbrica en la Educación Primaria en México

Luis Ángel Contreras Niño[1]
Universidad Autónoma de Baja California, México
Manuel González Montesinos
Universidad de Sonora, México
Erick Urías Luzanilla
Universidad Autónoma de Baja California, México

Compendio

Se presentan resultados de un estudio evaluativo cuyo objetivo fue caracterizar mediante una rúbrica los escritos producidos por 3,151 niños y niñas mexicanos al egresar de la educación primaria en el estado de Baja California/MX. Se describe la rúbrica utilizada y los procedimientos para capacitar en su uso a 31 jueces especialistas en lenguaje, así como para estimar la confiabilidad de sus juicios sobre los escritos. Se describen también las puntuaciones, por nivel de logro, en cada uno de los rasgos de la escritura evaluados mediante la rúbrica. Además, se aportan evidencias relacionadas con la validez del examen que incluyó el ítem de ejecución que capturó las muestras de la escritura. Finalmente, se sugieren estudios cualitativos y cuantitativos que amplíen y profundicen los hallazgos encontrados en esta evaluación a gran escala.

Palabras clave: Evaluación de la escritura; Rúbrica; Educación básica.

Writing Evaluation Rubric in the Elementary Education in Mexico

Abstract

Results of an evaluation study to characterize using a rubric the writings produced by 3,151 Mexican children when they left primary education in the Baja California/MX state are presented. It is described the rubric used and the procedures to train 31 judges, specialists in language, in the use of the rubric, as well as the procedures to estimate the reliability of their judges about the writings. We also describe the achievement level in each trait of the writings evaluated using the rubric. We also present some evidences related with the validity of the exam in which was included the performance item that captured the writing samples. Finally, qualitative and quantitative studies are suggested to expand and deepen the discoveries found in this large scale evaluation.

Keywords: Writing evaluation; Rubric; Basic education.

El presente estudio forma parte de otro más amplio cuyo propósito es diseñar y validar un examen de español destinado a monitorear la calidad de los aprendizajes que logran en esta área los egresados de las escuelas primarias del estado de Baja California/MX, México (Contreras, 2000; Contreras & Backhoff, 2004). El examen está alineado con el currículo oficial, por lo cual explora el dominio que tienen los examinados de las habilidades y conocimientos correspondientes a los ejes de Lengua Hablada, Lengua Escrita, Recreación Literaria y Reflexión sobre la Lengua. Se trata de una prueba de referencia criterial de gran escala, estructurada en cuatro versiones de examen que contienen 44 reactivos de opción múltiple cada una de ellas, y que incluyen también una pregunta de ensayo que captura una muestra de la escritura de los examinados. El examen fue aplicado a fines de 2001, con el apoyo del Sistema Educativo Estatal del estado de Baja California/MX, a una muestra estatal de 3,151 niños y niñas que presentaban diversas condiciones personales, escolares y familiares.

El trabajo de análisis curricular del área de español, que sirvió de base para construir la prueba, deja ver con claridad el gran énfasis que se otorga a la lengua escrita en el currículum adoptado para la educación primaria por la Secretaría de Educación Pública ([SEP], 1993) en la década pasada. Lo anterior puede ser constatado si consideramos que, en la tabla de especificaciones del examen, de un total de 182 contenidos estructurados, 84 contenidos (el 46.1 % del total del área de español) pertenecen al eje de lengua escrita.

Por tratarse de una prueba de gran escala, las especificaciones de ítems se orientaron a producir reactivos de opción múltiple. Sin embargo, el enfoque del

[1] Dirección: Universidad Autónoma de Baja California, Instituto de Investigación y Desarrollo Educativo, A. P. 453, Ensenada, BC, México, C.P. 22800. *E-mail:* angel@uabc.mx.

curriculum del área de español enfatiza en gran medida el desarrollo de habilidades comunicativas de producción, principalmente en el eje de lengua escrita, que requieren para su evaluación apropiada de ítems de respuesta construida, como los de ejecución. Aunque el comité que diseñó las especificaciones de ítems de la prueba, hizo un esfuerzo considerable para que al evaluar habilidades pudiera seleccionarse alguna dimensión de ellas cuya medición con ítems de opción múltiple aportara información relevante sobre su dominio, ello no fue posible en todos los casos; por lo que las especificaciones tienden a ignorar esas partes del curriculum, a pesar de ser importantes. Una excepción a lo anterior, fue la decisión de incorporar al examen una especificación para producir ítems de ejecución orientados a capturar una muestra de la habilidad para redactar de los estudiantes, correspondiente al eje de Lengua Escrita, y que se consideró demasiado importante en el contexto del currículo como para ser ignorada en el examen. En consecuencia, los objetivos específicos de este trabajo son dar cuenta tanto de las características que tienen los escritos que produjeron los niños en respuesta a los ítems de ejecución, así como de los procedimientos utilizados para efectuar dicha caracterización.

Aspectos Psicométricos de la Evaluación de la Escritura

En la actualidad existen numerosos procedimientos para evaluar el aprendizaje del lenguaje. Entre los más importantes figuran la observación directa de la ejecución del examinado en situaciones de aprendizaje auténticas, las listas de cotejo que orientan la observación en áreas específicas del lenguaje, los registros de muestras independientes de lectura o escritura que permiten observar su desarrollo, las iteraciones o repeticiones para evaluar la construcción de significado, las autoevaluaciones de estudiantes para determinar sus percepciones acerca de tópicos específicos, las entrevistas de proceso para observar las estrategias metacognitivas del alumno, la toma de muestras de escritura en ocasiones diferentes y sobre temáticas diferentes, los círculos literarios para evaluar la construcción de significado e integrar la instrucción y la evaluación, los inventarios de intereses para planear actividades y materiales, el análisis de errores inadvertidos para evaluar estrategias de decodificación y la evaluación de ejecuciones para juzgar la aplicación de estrategias, habilidades y conocimientos e integrar la evaluación con la instrucción, entre otros (*Cooper, 1997*).

En nuestro idioma, destaca la propuesta que formulan Bazán, Sánchez, Corral-Verdugo y Castañeda (2006) para emplear un sistema analítico, mediante ecuaciones estructurales, útil al estudio del aprendizaje y dominio de la lengua escrita, mismo que puede:

> ayudar a describir y explicar los patrones de relaciones de los resultados de los niños en pruebas de evaluación de desempeño, con otras mediciones, con la medición de otros rasgos o características del niño, y con el uso de distintos criterios de medición, lo cual se establece a partir de datos de diferencias individuales. (p. 96)

Tales procedimientos tienen en común que se refieren a la evaluación de ejecuciones, las cuales son definidas, según los principales estándares psicométricos para la evaluación psicológica y educativa, como medidas basadas en conductas o en productos, las cuales se establecen para emular contextos o condiciones de la vida real en las cuales se aplican conocimientos o habilidades específicos (American Educational Research Association, American Psychological Association, &National Council on Measurement in Education [AERA, APA, & NCME], 1999).

Dichos procedimientos han tenido su principal aplicación en el ámbito del salón, para investigar el aprendizaje, retroalimentar la instrucción o fundamentar la promoción de grado. Sin embargo, en la última década algunos de ellos han venido utilizándose con éxito en el contexto de la evaluación a gran escala, complementando a las tradicionales pruebas estandarizadas de opción múltiple que cumplieron esa función durante mucho tiempo. Entre ellos, destaca la evaluación de la escritura mediante una matriz de evaluación, también llamada confirmación de respuestas o rúbrica, como aquí se le denominará en adelante.

Se trata de una técnica para evaluar la construcción de significados mediante la escritura, que proporciona una vía para juzgarla mediante el examen de una muestra completa. Una rúbrica es una escala que describe uno o varios tipos y niveles de habilidad respecto a una ejecución determinada. Para ello, proporciona una guía para calificar que diferencia, en una escala articulada, entre un grupo de muestras producidas por estudiantes que responden la misma tarea evaluativa y al hacerlo sus ejecuciones se dispersan en un rango que abarca desde una respuesta excelente, hasta una que resulta inapropiada y necesita mejorar (University System of Georgia [USG], 2002).

Toda rúbrica se enfoca en la medición de cierto constructo, utiliza un rango para ordenar la ejecución y contiene características de actuación específicas arregladas en niveles que indican el dominio de un estándar.

A diferencia de los ítems de opción múltiple, que se enfocan en habilidades incrementales y son una vía indirecta para evaluar dimensiones discretas de habilidades comunicativas tales como la ortografía, gra-

mática y el uso de otras convenciones lingüísticas, la evaluación mediante rúbrica es una aproximación directa a la evaluación de la escritura, en la cual se pide al estudiante que escriba uno o más ensayos que son evaluados mediante algún método de calificación como el analítico, holístico o el de la habilidad primaria. La mayoría de los ensayos son escritos por demanda en respuesta inmediata a un instigador (*prompt*) verbal.

En el contexto de la evaluación del lenguaje se considera que la buena escritura es algo más que el empleo correcto de la gramática o el uso adecuado de la estructura de las oraciones o párrafos. Tales elementos de una muestra de escritura son importantes, pero lo es más la necesidad del autor de comunicar información, pensamientos e ideas. Ese es el verdadero propósito de la escritura (Swift, 2003).

En consecuencia, se requiere de una estrategia evaluativa que dé cuenta de la capacidad del estudiante para escribir, que lo ayude a analizar y mejorar su escritura mediante estándares de ejecución específicos, así como que proporcione un lenguaje común a planeadores, autoridades educativas, supervisores escolares, profesores, padres de familia y a los propios estudiantes, de manera que se mejore la comprensión y se reduzcan posibles confusiones al discutir y evaluar una muestra de la escritura de los alumnos.

En este sentido, el uso de la rúbrica en el contexto del salón y en el de la gran escala presenta ventajes importantes, entre las que destacan la promoción de expectativas consensuadas, conocimiento previo por parte del estudiante de los criterios para calificar su ejecución, reducción de la subjetividad del evaluador y la promoción de su consistencia al calificar, retroalimentación de la efectividad de la instrucción, promoción de estándares contra los cuales medir y documentar el progreso, apoyo a la rendición de cuentas y un enfoque integral del contenido, la ejecución, los estándares de ejecución y el trabajo del estudiante.

Existen dos tipos generales de rúbrica:

La Holística. Se emplea para medir el efecto global que produce en un evaluador una tarea evaluativa, con una guía diseñada para obtener una impresión general. Este tipo de rúbrica constituye una aproximación cualitativa, rápida y eficiente para juzgar una ejecución; sin embargo, no es una medida apropiada cuando se requiere evaluar qué tan bien el estudiante aplicó habilidades o conceptos específicos.

La Analítica. Se utiliza para calificar una tarea evaluativa mediante la asignación de puntajes en cada uno de los rasgos o elementos que la constituyen. Resulta apropiada cuando se desea comparar a estudiantes en cuanto a sus conocimientos, habilidades o la aplicación de ellos, contra estándares de excelencia. Es cuantitativa, subjetiva y puede no ser apropiada para algunas tareas evaluativas.

Un aspecto crítico del uso de una rúbrica es el entrenamiento de los evaluadores. Las personas involucradas en el proceso de calificación, deben participar en sesiones de entrenamiento que involucran una detallada discusión de la rúbrica y realizar sesiones de ensayo, en las que asignen a los escritos que sirven de prueba un porcentaje mínimo de puntajes exactos de acuerdo con los previamente establecidos por los coordinadores del examen. Los evaluadores que resultan incapaces de obtener los estándares correspondientes, en cuanto a lograr una comprensión cabal de la rúbrica y los criterios para asignar calificaciones, tras repetidos intentos y una realimentación adecuada, no participan en la evaluación real.

Método

En esta sección se describen los participantes en el estudio, los materiales utilizados y los procedimientos que se emplearon para diseñar una rúbrica analítica para evaluar las respuestas a las preguntas de ensayo del examen de español. También se describen los procedimientos empleados para entrenar a evaluadores en el uso apropiado de la rúbrica y para garantizar la confiabilidad de sus juicios al calificar los ensayos de los examinados. Además, se describen los procedimientos que se siguieron para analizar la información derivada de la aplicación de la rúbrica, así como para contrastar los resultados obtenidos al utilizar la rúbrica con aquellos que lograron los niños y niñas en los reactivos de opción múltiple del examen, mismos que exploraron su dominio de los contenidos correspondientes a las líneas de formación que contempla el eje de Lengua Escrita del área de español.

Participantes

Como ya se mencionó, el examen fue aplicado a fines de 2001 a una muestra estatal estratificada de examinados conformada por 3,151 niños y niñas que egresaron en el ciclo escolar 2000-2001, de 48 escuelas de educación primaria en Baja California/MX, cuando ingresaron a la escuela secundaria. En conjunto, los examinados presentaban diversas condiciones en cuanto a municipio y delegación donde estudiaron, tipo de control de la escuela, turno escolar al que asistieron, nivel socioeconómico de la familia y otras variables contextuales más que fueron exploradas.

También participaron en el estudio profesores, egresados, y estudiantes avanzados de carreras relacionadas con el lenguaje, a quienes se habilitó como evaluadores.

Materiales y Procedimientos

Los principales materiales que fueron empleados en el estudio, incluyen la especificación de reactivos y los ítems mediante los cuales se capturó la muestra de la escritura de los examinados, la rúbrica que se utilizó para calificar los ensayos producidos por los niños y niñas que fueron evaluados y los materiales que se emplearon para capacitar a los evaluadores que emplearon la rúbrica para calificar los ensayos.

La Tabla 1 presenta la especificación que produjo los cuatro ítems de ensayo, misma que también ilustra uno de ellos.

Puesto que una evaluación de ejecución no tiene una clave de respuesta correcta o incorrecta absoluta, se requieren reglas más detalladas para calificar. Así, se requería una rúbrica para evaluar los ensayos de los examinados. Para ello, se consideraron varias rúbricas existentes para evaluar la escritura a fin de encontrar la más apropiada para explorar el dominio de la habilidad de redacción, según lo establece el contenido que aparece en el programa de estudios de 5º grado (véase la descripción del contenido a evaluar en la Tabla 1).

Al respecto, se consideró que dadas las características de los escritos producidos por los niños y niñas como su brevedad, respuesta a un instigador, dominio previo del contenido y otras relacionadas, podría utilizarse una rúbrica desarrollada dentro del modelo de escritura de 6+1 rasgos, propuesto a principios de los años ochenta del siglo pasado por investigadores del laboratorio psicométrico regional del noroeste de los Estados Unidos (Kozlow & Bellamy, 2005; Northwest Regional Educational Laboratory [NWREL], 2001). De manera específica, se tradujo y adaptó la rúbrica desarrollada para evaluar a nivel estatal la escritura en los grados K3 a K5 (Wolfe, Dalton, & Neuburger, 1993), mismos entre los que se encuentra el que corresponde al egreso de la educación primaria en México. La Tabla 2 presenta una breve descripción de los rasgos de la rúbrica que se utilizó para evaluar los escritos de los examinados que respondieron las preguntas de ensayo.

Por su parte, cada uno de los seis rasgos fue evaluado en una escala de seis niveles de ejecución, correspondiendo el 6 a la mejor ejecución y el 1 a la más deficiente. En consecuencia, la calificación máxima posible fue 36 puntos y la mínima 6.

Con el propósito de ilustrar la escala de calificaciones que se utilizó al medir cada rasgo de la escritura considerado en la rúbrica, en la Tabla 3 se muestran los niveles de ejecución que fueron considerados por los jueces para evaluar el primero de los rasgos de la escritura de los niños; es decir, el que se refiere a las ideas y contenido incluidos en los escritos producidos por los examinados.

Tabla 1
Especificación para Producir Ítems que Capturan una Muestra de la Escritura de los Examinados

Contenido: Redacción individual y colectiva de textos considerando título, secuencia y relación entre ideas, atendiendo a la exposición de relaciones causales (5º grado).

Se espera que un logro fundamental en el área de español y en particular en el eje de Lengua Escrita, sea el desarrollo de la capacidad para escribir textos que presenten una mínima coherencia. Por ello, se requiere de un ítem que explore la habilidad del niño para redactar un texto breve y el cual presente las siguientes características:

· Será un ensayo dirigido; es decir, el ítem solicitará al alumno redactar aspectos específicos de un tema particular, a fin de evitar divagaciones e información poco relevante.

· El instigador verbal solicitará al estudiante escribir sobre un tema cuyo contenido domine, que le resulte atractivo, que esté acorde con su nivel de conceptualizaciones (se trata de un niño o niña de alrededor de 12 años) y que no revele aspectos íntimos de su personalidad o de su familia. Por ejemplo, puede solicitar que escriba sobre sus pasatiempos, amistades, etc.

· El instigador verbal solicitará al estudiante fundamentar su punto de vista y redactar al menos 10 renglones.

Ejemplo:
Escribe una composición en la que describas a tu mejor amigo o amiga. Piensa que tu lector es una persona que no sabe nada de ti; piensa entonces en qué aspectos debes incluir en tu composición para que tu lector tenga una idea clara de cómo es tu mejor amigo o amiga y porqué consideras que lo es. Utiliza oraciones completas, y no descuides la puntuación y la ortografía. Una vez que termines tu composición, léela y corrige tus errores. *Escribe un mínimo de 10 renglones.*

Tabla 2
Descripción Resumida de los Rasgos que Evalúa la Rúbrica del Examen de Español

Rasgo de la escritura evaluado	*Descripción resumida*
Ideas y contenido	*Explicar el tema o mensaje.* Este rasgo establece el tema del escritor, el foco de su mensaje, junto con los detalles de apoyo que desarrollan y enriquecen dicho tema. Las ideas principales se comunican y apoyan mediante detalles informativos que muestran una exploración del tema apropiada para la audiencia y propósito comunicativo.
Organización	*Planear y emplear conexiones claras de principio a fin.* Este atributo se refiere a la estructura interna del escrito, la cual incluye el hilo conductor del mensaje central y los patrones que mantienen unificado al escrito. La estructura organizativa puede estar basada en la comparación y el contraste, la cronología de un evento u otro patrón identificable. Cuando la organización es apropiada, el escrito crea en el lector un sentido de anticipación. Los eventos se suceden lógicamente, la información se dosifica de manera que el lector nunca pierde el interés o el panorama de lo que el escritor pretende. Las transiciones mueven al lector de un punto al siguiente.
Voz	*Proyectarse como una persona real.* La voz es la manifestación del escritor a través de las palabras, su sello personal; el sentido de que una persona real se dirige a nosotros y de que tiene interés en el mensaje. Es el corazón y el alma del escrito; su magia, su humor, su sentimiento, su vida. Este rasgo muestra el interés y compromiso del escritor con el tema. La voz variará de acuerdo con el propósito y el tipo de texto, pero debe ser apropiadamente formal o casual, distante o íntima, dependiendo del propósito o la audiencia.
Elección de palabras	*Elegir con cuidado palabras para formar una imagen en la mente del lector.* Este rasgo refleja el uso específico que hace el escritor de las palabras y oraciones, a fin de transportar el mensaje de manera interesante, precisa y natural; apropiada para la audiencia y el propósito. La elección de palabras no solo sirve para comunicar de manera funcional, sino que mueve al lector hacia una nueva visión de las cosas e ilumina y expande las ideas. La elección correcta de palabras se caracteriza por la habilidad de emplear con precisión palabras de uso cotidiano.
Fluidez de las oraciones	*Crear oraciones que hagan sentido y que luzcan articuladas cuando se leen en voz alta.* Este rasgo da cuenta del ritmo y flujo del lenguaje, del sonido de los patrones de palabras y de la variedad de estructuras de las oraciones. ¿Cómo suena el escrito si se lee en voz alta? Esa es la prueba a realizar. Un escrito fluido tiene cadencia, poder, ritmo y movimiento.
Convenciones	*Utilizar de manera correcta la ortografía, puntuación, escritura de párrafos y demás reglas del español.* Este atributo se refiere a la mecánica del escrito, al apego a las convenciones de la lengua. Un escrito apropiado, usualmente ha sido revisado y editado con cuidado. La caligrafía y limpieza no se califican como parte de este rasgo.

Selección y Entrenamiento de los Evaluadores

Las características de la evaluación de la ejecución a gran escala mediante ítems de respuesta construida hacen que resulte compleja, tardada y costosa. Los principales requerimientos para ello incluyen el desarrollo de matrices de evaluación o rúbricas, la selección y entrenamiento de jueces expertos, y la operación de prolongadas sesiones de evaluación. En consecuencia, tras su aplicación a gran escala, los ensayos de los niños no pudieron ser evaluados de inmediato por falta de recursos humanos y financieros, sino hasta recientemente.

Para ello, se hizo una convocatoria a egresados y estudiantes avanzados de la carrera Lengua y Literatura de Hispanoamérica, del campus Tijuana/MX de la Universidad Autónoma de Baja California, y a profesores de español de la Escuela Normal Estatal de Ensenada, Baja California/MX, a quienes se dio un entrenamiento específico en el manejo de matrices de evaluación de la escritura, particularmente en cuanto al uso de la rúbrica desarrollada. La capacitación se hizo en dos sesiones e incluyó también el entrenamiento de dichos especialistas para lograr estabilidad de sus observaciones y juicios, a fin de

Tabla 3
Escala de Ejecución para Evaluar en los Escritos el Dominio del Rasgo Ideas y Contenido

Ideas y Contenido *Explicar el tema o mensaje*		
6 *El escrito es muy claro, resulta interesante y está bien enfocado. Mantiene la atención del lector de principio a fin.* - El escritor tiene un excelente dominio del tema y ha seleccionado cuidadosamente detalles que explican con claridad las ideas principales - Destacan las ideas principales y los detalles de apoyo - El escritor seleccionó contenido y detalles apropiados para el propósito y la audiencia - El escritor establece conexiones y comparte sus hallazgos	*5* *El escrito resulta claro, interesante y bien enfocado. Mantiene la atención del lector.* - El escritor tiene dominio del tema y eligió con cuidado los detalles que explican con claridad las ideas centrales. - El lector puede identificar fácilmente las ideas principales y los detalles de apoyo. - El escritor empató la forma en que presenta el tema con el propósito y la audiencia. - El escritor establece conexiones y comparte sus hallazgos	*4* *El escrito es claro y se apega al tema. Mantiene la atención del lector.* - El escritor muestra conocimiento del tema y eligió detalles que ayudan a explicar la idea principal. - El lector puede identificar la idea principal y detalles de apoyo. - El lector percibe que el escritor está conciente del propósito y la audiencia. - El escritor establece algunas conexiones y pueden estar presentes nuevos hallazgos
3 *El lector puede entender lo que el escritor trata de decir, pero el escrito no mantiene la atención del lector de principio a fin.* - El escritor tiene cierto dominio del tema; algunas ideas pueden ser claras, mientras que otras no lo son o no parecen encajar. - El escrito no cuenta con suficientes detalles, son demasiado generales o no tienen relación con las ideas. - El lector ubica algunas formas en que el escrito empata con el propósito y la audiencia, pero no siempre resulta claro. - El escritor establece conexiones obvias y predecibles.	*2* *El escrito no resulta muy claro y presenta pocos detalles apropiados.* - El escritor tiene poco dominio del tema; las ideas no resultan claras. - El escrito presenta detalles limitados, que se repiten o que no tienen relación con las ideas. - El lector no está seguro del propósito ni de la idea central del escrito, pero puede hacer algunas suposiciones al respecto.	*1* *El escrito no es claro y pareciera no tener propósito.* - Las ideas del escritor son muy limitadas o pueden dispersarse en diversas direcciones. - Es difícil entender lo que el escritor realmente quiso decir.

garantizar índices aceptables de confiabilidad en sus calificaciones. Además, se proporcionó una preparación en el manejo de procedimientos automatizados para la captura de sus juicios, mediante una base de datos que operó en línea desde una página Web que se diseño para tal efecto (Ver Figura 1). Así, 31 jueces calificaron cada uno de ellos 100 ensayos durante aproximadamente un mes.

Resultados y Discusión

En esta sección se presentan los principales resultados obtenidos durante el proceso de calificar los ensayos por parte de los jueces, los obtenidos por los niños y niñas en los ensayos calificados mediante la rúbrica y los alcanzados por ellos en las demás partes del examen que exploraron el dominio de los contenidos de lengua escrita.

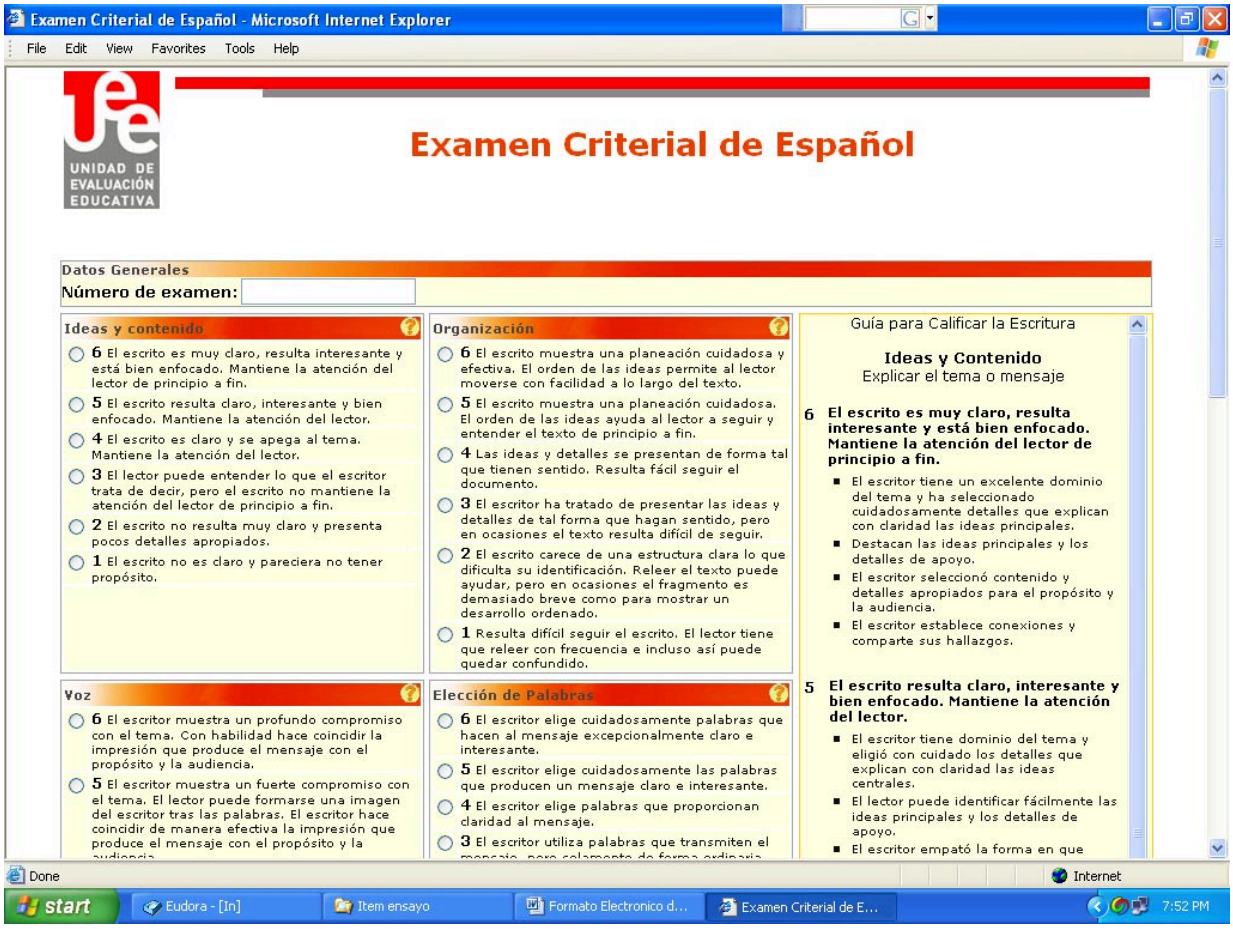

Figura 1. Sección de una pantalla que captura en línea las evaluaciones de los jueces

Análisis de la Confiabilidad de los Juicios Emitidos por los Evaluadores

La interpretación válida de los puntajes en la rúbrica depende sustancialmente de la confiabilidad de los jueces que calificaron las ejecuciones. La confiabilidad de los jueces es una propiedad del proceso de medición y como tal debe ser determinada en cada ejercicio particular del proceso. Para estimar la confiabilidad se identificaron tres aproximaciones psicométricas aplicables en este caso.

Aproximación a la Confiabilidad Basada en la Consistencia entre Jueces. En esta aproximación se toma como base el cálculo el Coeficiente α de Cronbach como una medida del grado en que los puntajes asignados por jueces múltiples convergen para medir un constructo en común. El coeficiente α es una medida de consistencia interna de las calificaciones. Si el estimado resulta bajo, implica que la variabilidad en los puntajes combinados se debe a variabilidad atribuible al error aleatorio y no al puntaje verdadero en el constructo de interés (Hatcher, 1998).

Para este caso, se desagregó la base de datos del ítem 45 de la prueba para analizar los puntajes de los seis rasgos en cada versión del examen por separado. Así, los coeficientes correspondientes a los puntajes otorgados por los jueces se muestran en la Tabla 4.

Tabla 4
Coeficientes de Consistencia Interna de las Calificaciones Otorgadas por los Jueces

Modelo	Alfa
1 (n=756)	.935
2 (n=747)	.927
3 (n=702)	.932
4 (n=671)	.909

Como los tamaños originales de muestra son considerables, se seleccionó una muestra aleatoria de 50 casos de cada modelo y se repitió el cálculo para descartar que los coeficientes se inflen artificialmente debido al

tamaño de muestra en cada forma o versión del examen. Los resultados con 50 casos se muestran en la Tabla 5.

Tabla 5
Coeficientes de Consistencia Interna de las Calificaciones con n = 50

Modelo	Alfa
1	.958
2	.923
3	.946
4	.943

Como puede observase en ambos casos los coeficientes alfa son superiores a .90, lo que definitivamente constituye una razón sólida para concluir que las calificaciones otorgadas por los jueces son consistentes y convergen en un constructo común.

Aproximación a la Confiabilidad Basada en el Proceso de Medición. Otra aproximación que se empleó para estimar la confiabilidad entre jueces se basa en la información que proporciona el procedimiento de medición en su conjunto. Esta aproximación se implementa realizando un análisis de componentes principales (ACP) sobre los datos de cada escala. Este método es de aplicación óptima cuando la escala que se analiza se ha diseñado para medir un sólo constructo unidimensional (e.g. competencia en escritura).

Los puntajes otorgados por los jueces se someten al ACP para determinar la cantidad de varianza compartida que se puede atribuir al primer componente extraído. El porcentaje de varianza explicada por el primer componente proporciona una indicación del grado en que los jueces están coincidiendo. Si el porcentaje de varianza es alto (e.g. 60%), se tiene también una indicación de que los jueces están valorando un constructo unidimensional (Stemler, 2004).

Se aplicó el procedimiento sobre las bases desagregadas por forma/versión del examen y se obtuvieron los resultados que se muestran a continuación en la Tabla 6 compuesta.

Como puede verse en la Tabla compuesta, el procedimiento ACP extrajo sólo un componente principal con valor *eigen* superior a 1. En todos los casos el porcentaje de varianza explicada por el componente es superior al 60%.

La ventaja de esta aproximación es que permite asignar puntajes finales a los examinados con base en la dimensión de mayor peso; es decir, el primer componente principal. La desventaja es que se asume que los puntajes de los jueces están exentos de error de medición.

Aproximación a la Confiabilidad de los Jueces Basada en el Modelo de Rasch. El modelo Rasch de facetas múltiples fue propuesto por Linacre (1994), para analizar de forma exhaustiva la contribución individual de cada faceta y de cada elemento en un proceso de medición. En el caso particular del ítem de ensayo las tres facetas son: los jueces, los examinados y los ítems de la escala para cada rasgo. Específicamente, el modelo permite estimar la severidad de los jueces, la habilidad de los examinados y la dificultad de los ítems. Con propósitos de ilustración, la Figura 2 muestra la salida de resultados del análisis efectuado con el programa de cómputo denominado FACETS (Linacre, 1994), que compara la severidad de los jueces en la versión 1 de la prueba.

Los estadígrafos de ajuste interno (*Infit*) y externo (*Outfit*) indican el grado en que el comportamiento de cada juez se ajusta a las expectativas del modelo Rasch, tomando en cuenta la calibración individual de severidad obtenida para cada juez. Por regla general los valores *Infit* y *Outfit*, en Mean Square (MsSq), deben mantenerse dentro del intervalo .80 a 1.30. Para los valores estandarizados (ZStd) el rango aceptable es de -2 a +2.

La aproximación vía FACETS permite además obtener una estimación de la consistencia intra-juez. En particular, los valores *Infit* y *Outfit* proporcionan una medida del grado en que las calificaciones otorgadas por los jueces son internamente consistentes, al mantenerse dentro de la expectativa que el modelo Rasch crea para cada juez, dado su patrón observado de calificaciones emitidas dentro del proceso en su conjunto. Valores de *Infit* MsSq mayores a 1.30 indican mayor variabilidad intra-juez que la que se esperaría con base al modelo (Stemler, 2004).

Esta aproximación tiene la ventaja de no requerir que todos los jueces califiquen todos los ítems para lograr una estimación de la confiabilidad entre jueces, como fue el caso al calificar los escritos producidos en respuesta a los ítems de la prueba. En lugar de ello, los jueces pueden calificar un subconjunto particular de ítems y, mientras exista suficiente conectividad (Linacre, 1994; Linacre, Englehard, Tatum, & Myford, 1994) entre jueces y calificaciones, será posible comparar directamente a los jueces.

La Figura 2 presenta un reporte de la medición de los jueces; esto es, un indicador del nivel de severidad de cada juez en lo individual, junto con varios estadísticos de ajuste que ayudan a diagnosticar en qué medida cada juez fue consistente con su propio uso de la rúbrica para evaluar los ensayos. La utilidad de la información en la figura radica en que puede ser comparada simultáneamente la severidad relativa de todos los jueces. Así, los índices de severidad de los jueces (*Measure*) son útiles para estimar en qué medida existen diferencias sistemáticas entre jueces respecto a su nivel de severidad y, en su caso, proceder a ajustarlas.

Tabla 6
Análisis de Componentes Principales de las Puntuaciones Otorgadas por los Jueces

Versión 1

Total de varianza explicada

Componente	Valores *eigen* iniciales			Sumas extraídas de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	4.545	75.747	75.747	*4.545*	*75.747*	*75.747*
2	.553	9.222	84.969			
3	.272	4.541	89.510			
4	.239	3.982	93.492			

Versión 2

Total de varianza explicada

Componente	Valores *eigen* iniciales			Sumas extraídas de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	4.427	73.778	73.778	*4.427*	*73.778*	*73.778*
2	.556	9.273	83.051			
3	.345	5.753	88.805			
4	.272	4.534	93.339			

Versión 3

Total de varianza explicada

Componente	Valores *eigen* iniciales			Sumas extraídas de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	4.485	74.752	74.752	*4.485*	*74.752*	*74.752*
2	.724	12.074	86.826			
3	.254	4.234	91.061			
4	.214	3.568	94.629			

Versión 4

Total de varianza explicada

Componente	Valores *eigen* iniciales			Sumas extraídas de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	4.167	69.456	69.456	*4.167*	*69.456*	*69.456*
2	.680	11.338	80.795			
3	.354	5.905	86.700			
4	.314	5.234	91.933			

Nota. En los cuatro casos el método de extracción fue el análisis de componentes principales.

```
"Examen Espanol I45 Modelo1"  04-26-2007 13:30:06
Table 7.3.1  Item Measurement Report  (arranged by mN).

-------------------------------------------------------------------------------------------------------
|  Obsvd  Obsvd  Obsvd  Fair-M|        Model | Infit      Outfit    |Estim.|       |                    |
|  Score  Count Average Avrage|Measure  S.E. | MnSq ZStd  MnSq ZStd|Discrm| PtBis | N Item             |
-------------------------------------------------------------------------------------------------------
|    301    114     2.6   2.53|    .89   .15 | 1.71  4.1  1.65  3.8|  .37 |   .34 | 6 Convenciones     |
|    327    114     2.9   2.75|    .32   .15 |  .89  -.8   .82 -1.3| 1.15 |   .58 | 4 Palabras         |
|    345    114     3.0   2.89|   -.05   .14 |  .66 -2.7   .71 -2.3| 1.28 |   .64 | 5 Fluidez          |
|    352    114     3.1   2.95|   -.19   .14 |  .86 -1.0   .90  -.7| 1.10 |   .60 | 2 Organizacion     |
|    363    114     3.2   3.03|   -.41   .14 |  .79 -1.5   .83 -1.2| 1.15 |   .59 | 1 Ideas            |
|    371    114     3.3   3.10|   -.56   .14 |  .84 -1.2   .85 -1.1| 1.15 |   .62 | 3 Voz              |
-------------------------------------------------------------------------------------------------------
|  343.2  114.0   3.0   2.87|    .00   .14 |  .96  -.5   .96  -.5|      |   .56 | Mean (Count: 6)    |
|   23.4    .0    .2    .19|    .48   .00 |  .34  2.2   .31  2.0|      |   .10 | S.D. (Populn)      |
|   25.6    .0    .2    .21|    .53   .01 |  .38  2.4   .34  2.2|      |   .11 | S.D. (Sample)      |
-------------------------------------------------------------------------------------------------------
Model, Populn: RMSE .14  Adj (True) S.D. .46  Separation 3.24  Reliability .91
Model, Sample: RMSE .14  Adj (True) S.D. .51  Separation 3.57  Reliability .93
Model, Fixed (all same) chi-square: 65.5  d.f.: 5  significance (probability): .00
Model,  Random (normal) chi-square: 4.7  d.f.: 4  significance (probability): .32
```

Figura 2. Estimación de la severidad de los jueces mediante el modelo Rasch de facetas múltiples (Salida de FACETS)

Con las estimaciones sobre la confiabilidad de los jueces mediante los métodos antes descritos fue posible determinar los grados de severidad o laxitud de los jueces y sobre todo verificar que estas diferencias se conservaban dentro de un límite previsible en cada versión de la prueba, lo que permitió concluir que los resultados obtenidos de la evaluación de los escritos son interpretables.

Características Generales de la Escritura de los Egresados de la Educación Primaria en Baja California/MX

En consecuencia, se procedió con confianza a determinar las puntuaciones generales otorgadas por los jueces mediante la rúbrica, al total de los ensayos. Estos resultados se presentan en la Tabla 7.

Por su parte, las calificaciones que otorgaron los jueces a los ensayos producidos en respuesta al ítem, correspondientes a cada versión de la prueba, se muestran en la Tabla 8.

De conformidad con la rúbrica, el conjunto de los jueces que calificaron los escritos que produjeron los examinados en respuesta a los cuatro ítems que instigaban su redacción de al menos 10 renglones sobre un tema familiar a ellos, asignaron al conjunto de ensayos las siguientes puntuaciones totales (Ver Tabla 9 y Figura 3)

Tabla 7
Puntuaciones de los Jueces Otorgadas al Total de los Ensayos en Cada Rasgo de la Rúbrica

Estadístico	Ideas	Organización	Voz	Palabras	Fluidez	Convenciones
Media	3.3518	3.1965	3.3545	3.0695	3.1139	2.8907
Desviación estándar	1.2218	1.1455	1.2254	1.0873	1.0828	1.0513
N válido = 2993						

Tabla 8
Puntuaciones de los Jueces Otorgadas en Cada Rasgo de la Rúbrica, según la Versión de la Prueba

Modelo	Estadístico	Ideas	Organización	Voz	Palabras	Fluidez	Convenciones
1	Media	3.36	3.24	3.32	3.09	3.15	2.98
	Desviación estándar	1.27	1.22	1.24	1.16	1.14	1.10
	N válido=756						
2	Media	3.45	3.34	3.53	3.22	3.26	3.11
	Desviación estándar	1.30	1.14	1.30	1.11	1.06	1.04
	N válido=747						
3	Media	3.37	3.22	3.43	3.06	3.09	2.78
	Desviación estándar	1.19	1.17	1.24	1.09	1.12	0.97
	N válido=702						
4	Media	3.23	2.97	3.13	2.90	2.96	2.68
	Desviación estándar	1.15	1.04	1.11	0.98	1.01	1.06
	N válido=671						

Puntuaciones Promedio en los Rasgos

Las puntuaciones promedio asignadas por los jueces a cada uno de los rasgos de la escritura de los estudiantes se muestran en la Tabla 10. Como puede observarse, casi todos los rasgos fueron evaluados con un puntaje promedio por encima del nivel 3 en una escala donde 1 es la calificación mínima y 6 la máxima. Una excepción fue el rasgo de la escritura denominado Apego a las convenciones de la lengua, que resultó ligeramente inferior. Los rasgos mejor evaluados fueron las ideas o contenido de los escritos y la voz del escritor, con puntuaciones promedio de 3.36.

Tabla 9
Puntuación Promedio en la Rúbrica del Total de los Ensayos

Promedio de calificación	19
Calificación Mínima	6
Calificación Máxima	36
Alumnos evaluados	2,993

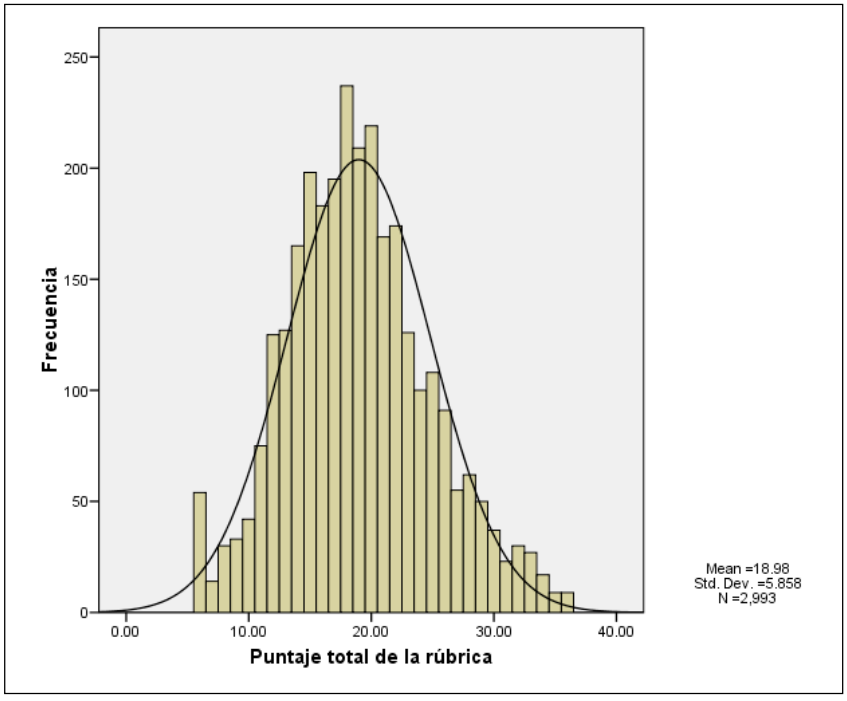

Figura 3. Distribución de las puntuaciones totales en la rúbrica

Tabla 10
Puntuaciones Promedio Asignadas a los Rasgos, según la Rúbrica

Estadístico	Ideas	Organización	Voz	Palabras	Fluidez	Convenciones
Puntaje promedio	3.36	3.20	3.36	3.07	3.11	2.90
Puntaje mínimo	1.00	1.00	1.00	1.00	1.00	1.00
Puntaje máximo	6.00	6.00	6.00	6.00	6.00	6.00
N válido = 2,993						

Nivel de Dominio en Cada Rasgo de la Escritura

En opinión del conjunto de jueces, el nivel de dominio que mostraron los niños en cada rasgo de su escritura se muestra en la Tabla 11. Para cada rasgo se presenta la frecuencia y el porcentaje de examinados que fueron ubicados en cada uno de los niveles de dominio, según la rúbrica utilizada.

Tabla 11
Frecuencias y Porcentajes por Nivel de Dominio en Cada Rasgo de la Escritura Evaluado

	Rasgos de la escritura evaluados por los jueces mediante la rúbrica											
Nivel de dominio*	Ideas del escrito		Organización del escrito		Voz del escritor		Dominio de las palabras		Fluidez		Convenciones	
	Nº de casos	%	Nº de casos	%	Nº de casos	%	Nº de casos	%	Nº de casos	%	Nº de casos	%
1	178	05.9	168	5.6	182	6.1	152	5.1	178	5.9	246	8.2
2	533	17.8	666	22.3	560	18.7	779	26.0	644	21.5	841	28.1
3	1,008	33.7	1037	34.6	926	30.9	1147	38.3	1183	39.5	1119	37.4
4	763	25.5	743	24.8	803	26.8	600	20.0	707	23.6	605	20.2
5	354	11.8	290	9.7	384	12.8	254	8.5	209	7.0	145	4.8
6	157	05.2	89	3.0	138	4.6	61	2.0	72	2.4	37	1.2

Nota. * El 6 representa el nivel de dominio más alto y el 1 el más bajo.

Tabla 12
Relación entre Rasgos de la Escritura y Conocimientos y Habilidades Relacionados con la Escritura Evaluados (ítems de opción múltiple)

	Total de conocimientos y habilidades de escritura	Manejo de letras y sílabas: ortografía	Redacción y elaboración de resúmenes	Comprensión y creación de géneros populares	Creación oral y escrita de géneros populares	Desarrollo de vocabulario mediante campos semánticos	Calificación total obtenida en la rúbrica	Ideas	Organización	Voz	Palabras	Fluidez	Convenciones
Total de conocimientos y habilidades de escritura	1												
Manejo de letras y sílabas: ortografía	0.783[1]	1											
Redacción y elaboración de resúmenes	0.466[1]	0.173[3]	1										
Comprensión y creación de géneros populares	0.606[1]	0.262[3]	0.179	1									
Creación oral y escrita de géneros populares	0.600[1]	0.260[3]	0.106	0.195	1								
Desarrollo de vocabulario mediante campos semánticos	0.362[1]	0.200[3]	0.178	0.175*	0.047	1							
Calificación total obtenida en la rúbrica	0.335[1]	0.292[3]	0.178	0.159	0.179	0.133	1						
Ideas	0.273[3]	0.232[3]	0.135	0.127	0.165	0.104	0.882[1]	1					
Organización	0.293[3]	0.253[3]	0.170	0.141	0.144	0.120	0.900[1]	0.792[2]	1				
Voz	0.259[3]	0.221[3]	0.142	0.110	0.160	0.086	0.872[1]	0.784[2]	0.763[2]	1			
Palabras	0.264[3]	0.228[3]	0.131	0.126	0.147	0.109	0.879[1]	0.724[2]	0.732[2]	0.737[2]	1		
Fluidez	0.296[3]	0.252[3]	0.153	0.146	0.161	0.126	0.882[1]	0.694[2]	0.756[2]	0.682[2]	0.757[2]	1	
Convenciones	0.355[1]	0.329[3]	0.192	0.177	0.146	0.151	0.743[1]	0.523[2]	0.590[2]	0.495[2]	0.595[2]	0.687[2]	1

Nota.* $p < .05$; para el resto de las correlaciones $p < .01$.

Por su parte, la Tabla 12 relaciona la ejecución en la rúbrica con los conocimientos y habilidades cuyo dominio fueron explorados con ítems de opción múltiple.

A partir de la información contenida en la Tabla 12 se pueden formular las siguientes observaciones: Las celdas identificadas con [1] presentan las correlaciones más altas, mismas que se dieron entre el total de conocimientos y habilidades que fueron explorados con ítems de opción múltiple y cada una de las líneas de formación que lo integran, así como con la calificación total de la rúbrica y el rasgo de ésta denominado *Convenciones*. Lo mismo sucede con las correlaciones entre la calificación total en la rúbrica y los rasgos que la integran. Por su parte, las celdas identificadas con [2] presentan correlaciones moderadamente altas que se dan entre los propios rasgos de la rúbrica. En cuanto a las celdas identificadas con [3] presentan correlaciones moderadas entre el total de conocimientos y habilidades que fueron explorados con ítems de opción múltiple y todos los rasgos de la rúbrica, así como entre la ortografía relacionada con el manejo de letras y sílabas y las demás variables que aparecen en la matriz. Las relaciones mencionadas, así como las demás que resultaron bajas pero significativas, muestran que existe coherencia entre todos los aspectos relacionados con la escritura, como fueron evaluados en el examen de español.

Conclusiones y Sugerencias

La realización del presente estudio evaluativo permitió obtener resultados y experiencias importantes que permiten formular las siguientes conclusiones y recomendaciones:

Resulta claro que la evaluación a gran escala de la ejecución de los estudiantes, referida a la aplicación de conocimientos y habilidades de lenguaje como en el caso de la redacción, resulta posible y es más significativa si se lleva a cabo mediante procedimientos como la rúbrica descrita, en vez explorarlos sólo a través de los tradicionales ítems de opción múltiple. Aunque el uso de ítems de respuesta construida en la evaluación a gran escala resulta complejo, prolongado y costoso en términos de los recursos humanos y materiales involucrados, vale la pena en el contexto de una evaluación nacional o estatal como la que aquí se describe.

Los resultados muestran que la rúbrica seleccionada, traducida y adaptada se ajustó al nivel de competencia comunicativa de los examinados y fue útil para describir las características de su escritura cuando egresan de la educación primaria. Prueba de ello es que, con pequeñas diferencias, la ejecución de los niños y las niñas fue similar y estuvo en el rango en todos los rasgos de la escritura cuyo dominio explora la rúbrica.

Al describir el presente estudio se enfatizó en el asunto de la confiabilidad de los juicios de los evaluadores que calificaron los ensayos mediante la rúbrica debido a que, por problemas operativos, no fue posible conseguir suficientes especialistas que apoyaran la evaluación bajo las condiciones del estudio (tiempo disponible para la capacitación, número de ensayos a evaluar, etc.). Ello ocasionó que ningún escrito pudiera ser evaluado por dos o más jueces independientes para calcular la confiabilidad mediante su grado de acuerdo. En consecuencia, fue necesario explorar la literatura especializada en el tema de la confiabilidad hasta encontrar métodos apropiados para estimar la confiabilidad en esas condiciones. Así, fue posible obtener una estimación razonable al conjuntar el coeficiente α de Cronbach como una medida del grado en que los puntajes asignados por jueces múltiples convergen para medir un constructo en común; el análisis de componentes principales sobre los datos de cada escala cuando se ha diseñado para medir un constructo unidimensional, como la competencia en escritura, a fin de asignar puntajes finales a los examinados con base en la dimensión de mayor peso; y mediante el uso de la extensión del modelo de Rasch, denominada de facetas múltiples, que hace posible determinar los grados de severidad o laxitud intra-juez e inter-juez para verificar que estas diferencias se conserven dentro de un limite previsible. El uso de tal estrategia convergente permitió concluir que los resultados obtenidos al evaluar los escritos eran interpretables.

Considerando de manera global la ejecución de los examinados, sus puntuaciones en la rúbrica están ligeramente sesgadas hacia las calificaciones bajas. Así, los puntajes promedios en los rasgos evaluados en la rúbrica apenas rebasan o están cerca de la calificación 3 en cada escala. Además, las frecuencias y porcentajes de examinados por nivel de dominio en cada rasgo evaluado, también se centran en el nivel de dominio 3.

Al relacionar las calificaciones obtenidas por los examinados en la rúbrica con los aciertos en otras partes del examen que se refieren al dominio de la lengua escrita, se observan correlaciones significativas en todos los casos, y son muy altas (superiores a .74) entre el total en la rúbrica y los rasgos que la conforman, así como altas (superiores a .35) entre el total en conocimientos y habilidades en lengua escrita y las líneas de formación que están incluidas en dicha sub-área. Las demás relaciones resultaron bajas pero significativas. Estos resultados muestran que existe alta congruencia y coherencia en los aspectos involucrados en la escritura, como son evaluados en el examen, y constituyen evidencias de validez de la prueba.

Los resultados en este estudio se consideran preliminares. Por ejemplo, las relaciones entre los rasgos

de la rúbrica y otras líneas de formación evaluadas en el examen solo consideran las correlaciones entre sus puntuaciones. Por ello, se consideran necesarios estudios cualitativos y cuantitativos que saquen el máximo provecho a los cuantiosos datos recabados mediante esta evaluación a gran escala. En particular, se sugieren un estudio lingüístico de los escritos y otro para desarrollar un modelo causal como el que proponen Bazán et al. (2006), que relacione las puntuaciones en los rasgos de la rúbrica con los resultados en las líneas de formación relacionadas con la escritura que se exploran en el examen. También se sugiere replicar el estudio en otros países hispanoparlantes para comparar las características de los escritos que en ellos se producen.

Referencias

American Educational Research Association., American Psychological Association., & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

Bazán, A., Sánchez, B., Corral-Verdugo, V., & Castañeda, S. (2006). Utilidad de los modelos estructurales en el estudio de la lectura y la escritura. *Revista Interamericana de Psicología, 40*(1), 85-93.

Contreras, L. A. (2000). *Desarrollo y pilotaje de un examen de español para la educación primaria en Baja California/MX*. Unpublished master's dissertation, Instituto de investigación y Desarrollo Educativo, Universidad Autónoma de Baja California, Ensenada, México. Retrieved from http://eduweb.ens.uabc.mx/egresados/Tesis/indicetesis.htm

Contreras, L. A., & Backhoff, E. (2004). Metodología para elaborar exámenes criteriales alineados al currículo. In S. Castañeda (Ed.), *Educación aprendizaje y cognición, teoría en la práctica* (chap. 10). México, DF: Manual Moderno.

Cooper, J. D. (1997). Literacy: Helping children construct meaning (3nd ed.). Boston: Houghton Mifflin.

Hatcher, H. (1998). *Using the SAS System for factor analysis and structural equations modeling*. Cary, NC: The SAS Institute.

Kozlow, M., & Bellamy, P. (2005). *Research on the 6+1 trait writing model for improving student writing*. Retrieved February 22, 2007, from http://www.nwrel.org/ascd05/

Linacre, J. M. (1994). *Many-facet Rasch measurement*. Chicago: MESA.

Linacre, J. M., Engelhard, G., Tatum, D. S., & Myford, C. M. (1994). Measurement with judges: Many-faceted conjoint measurement. *International Journal of Educational Research, 21*(6), 569-577.

Northwest Regional Educational Laboratory. (2001). *Assessment: About 6+1 trait™ writing.* Retrieved December 18, 2007, from http://pareonline.net/getvn.asp?v=9&n=4

Secretaría de Educación Pública. (1993). *Educación básica. Primaria. Plan y programas de estudio.* México, DF: Author.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation, 9*(4). Retrieved from http://PAREonline.net/getvn.asp?v=9&n=4

Swift, B. (2003). *The mandate of a writing assistant*. Santa Cruz, CA: University of California. Retrieved October 22, 2007, from http://people.ucsc.edu/~davidlaw/Swift.html

University System of Georgia. (2002). *Assessment module 9. Post secondary model for integrating technology*. Retrieved January 03, 2003, from http://ci.colstate.edu/psit/An%20Assessment%20Overview.htm

Wolfe, B., Dalton, M., & Neuburger, W. (1993). *Oregon statewide writing assessment 1991 and 1992*. (ERIC Document Reproduction Service No. ED366960).

Received 13/11/2008
Accepted 02/03/2009

Luis Ángel Contreras Niño. Universidad Autónoma de Baja California, México.
Manuel González Montesinos. Universidad de Sonora, México.
Erick Urías Luzanilla. Universidad Autónoma de Baja California, México.