

Un modelo de pauta de corrección para ensayos argumentativos

Martha Jurado Salinas*

Universidad Nacional Autónoma de México
Centro de Enseñanza para Extranjeros

Elsa Viviana Oropeza Gracia**

Universidad Nacional Autónoma de México
Centro de Enseñanza de Lenguas Extranjeras

Resumen

Este trabajo tiene como objetivo describir las características de una pauta de corrección analítica para evaluar ensayos argumentativos producidos por estudiantes de nivel bachillerato. Fue aplicada a 4 041 pruebas escritas del *Examen de Habilidades Lingüísticas* (EXHALING). Se construyó con base en el constructo teórico plasmado en el Marco Curricular Común del Bachillerato. Se concluye que, debido a la complejidad que implica el diseño de rúbricas analíticas de este tipo, es conveniente aprovechar el producto y la experiencia para contribuir a desarrollar en los estudiantes una mayor conciencia de las demandas de la tarea, y para dotar a los profesores de un instrumento de retroalimentación focalizada.

Palabras clave: pautas de corrección, rúbricas analíticas, textos argumentativos, bachillerato.

Abstract

This paper aims to describe the characteristics of an analytic rating scale that was used to assess 4 041 argumentative essays produced by high school students. These essays are part of a wider research project whose objective was to design an exam (*Examen de Habilidades Lingüísticas*, EXHALING), in order to evaluate the linguistic abilities that high school students have when applying for a university within Mexico City. It follows the theoretical framework of the Marco Curricular Común del Bachillerato. After careful analysis of the complex process for developing this type of scales, we conclude that using the scale itself and sharing this experience might be useful to help students develop a better understanding of what it is expected from them when writing these types of essays and provide teachers with focalized feedback.

Keywords: rating scales, analytic rating scales, argumentative texts, high school.



IZTAPALAPA

Agua sobre lajas

* Doctora en Lingüística por el Colegio de México.
mjurado@unam.mx

** Maestra en Lingüística Aplicada por la Universidad Nacional Autónoma de México.
viviana.oropeza@cele.unam.mx

En 1962 la UNESCO divulgó las “Normas que deben aplicarse en materia de publicaciones científicas” (documento UNESCO / NS / 177). En dicho documento se reunieron los lineamientos generales para la redacción de publicaciones en diferentes lenguas que fueron socializadas, comentadas y de aceptación generalizada entre diversas comunidades científicas. Más de cinco décadas después, los resultados de diversas pruebas estandarizadas evidencian que la instrucción formal primaria y secundaria de nuestro país no ha cumplido a cabalidad con la tarea de capacitar a los estudiantes en la producción de textos escritos con un lenguaje claro y directo, con párrafos bien organizados y lingüísticamente adecuados; tampoco la educación preuniversitaria ni la universitaria han cumplido con capacitarlos para producir distintos tipos de escritos académicos, principalmente ensayos argumentativos y artículos científicos (Oropeza *et al.*, 2014). Estas habilidades lingüísticas constituirían la plataforma para que, en el nivel universitario, los estudiantes asumieran la escritura como un método de conocimiento inherente al trabajo científico de cualquier disciplina.

Aunque las instituciones han hecho esfuerzos por atender esta y otras carencias del sistema educativo nacional (asignaciones presupuestales cuantiosas, actualización de maestros, reformas y renovación de planes y programas de estudio, etc.), los logros han sido magros. Los documentos institucionales producto de estas iniciativas poseen un alto rigor metodológico; sin embargo, los profesores carecen de guías o instrumentos para hacerlos operativos en su práctica diaria, con la consecuente persistencia de deficiencias en la producción escrita de los estudiantes de todos los niveles.

Como parte de dichos esfuerzos, la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES) llevó a cabo un diagnóstico de las habilidades lingüísticas de los estudiantes que ingresan a las instituciones de educación superior (IES) del área metropolitana de la Ciudad de México. Para ello se diseñó el *Examen de Habilidades Lingüísticas* (EXHALING), cuya aplicación tuvo lugar en 11 universidades públicas y privadas del área metropolitana, durante el segundo semestre de 2011.

En este marco se diseñó una pauta de corrección que dota de operatividad a la categoría “Se expresa y comunica”, plasmada en el Marco Curricular Común del Bachillerato (MCC), el cual a su vez tiene como base teórica el concepto de *competencia comunicativa*. En un segundo nivel de documentación del constructo, la pauta se fundamenta en una teoría psicosociolingüística que enfoca los procesos de lectura y escritura desde un punto de vista cognitivo-discursivo. Desde esta perspectiva teórica, las habilidades de lectura y de escritura son concebidas como actividades contextualizadas social y culturalmente, que detonan en el lector la construcción de procesos cognitivos complejos.

Las pautas de corrección o *rúbricas* son un tipo de escalas de puntuación que se han utilizado ampliamente en el terreno de la evaluación educativa. Su construcción y validación han sido objeto de numerosos estudios desde las perspectivas tanto de la evaluación de procesos educativos inter-aula (Hamp-Lyons, 1990; Mertler, 2001; Moskal y Leydens, 2000; Dempsey *et al.*, 2009), como de la evaluación de productos de las pruebas de respuesta construida de distintos tipos de exámenes (diagnóstico, logro académico, dominio lingüístico, etc.), (Kozlow y Belami 2005; Lee, Yong-Won *et al.*, 2008; East, 2009; Knoch, 2009; Becker, 2011; Wind y Engelhard, 2013). En el aula, el uso de las rúbricas se aplica principalmente en el desarrollo de proyectos. Dentro de este contexto, dichos instrumentos son construidos colaborativamente por los grupos de trabajo. Por lo que respecta a la evaluación de productos, la bibliografía especializada da cuenta de la aplicación y validación de diversas pautas de corrección, pero se soslaya el hacer explícito su fundamento teórico y, aunque se menciona su valor como instrumento de retroalimentación, no se profundiza en las implicaciones que pueden tener para la concientización de la tarea y el consecuente desarrollo de la habilidad escrita de los estudiantes.

El proceso de construcción de la pauta de corrección del EXHALING evidenció, en primer lugar, la dificultad que implica el diseño de instrumentos tanto holísticos como analíticos válidos y confiables; en segundo lugar, la necesidad de crear instrumentos eclécticos (mezcla de rúbricas analíticas con listas de comprobación) para evaluar adecuadamente las distintas subcompetencias; y finalmente, la imposibilidad de evaluar, mediante un instrumento de esta naturaleza, aspectos procesuales como el componente estratégico del aprendizaje. Su aplicación a los 4 041 ensayos fue una tarea titánica; los objetivos no se cumplieron a cabalidad en tanto que no se pudieron evaluar todas las subcompetencias previstas; sin embargo, el producto y la experiencia ganada pueden ser aprovechados para que los estudiantes del bachillerato, mediante la autoevaluación, desarrollen una conciencia de los distintos componentes de la escritura ensayística, así como para que los profesores cuenten

con un instrumento relativamente estandarizado para una población amplia a fin de proporcionar a sus alumnos una retroalimentación focalizada de sus escritos.

Los objetivos de este estudio son:

- ✦ Describir las características de la pauta de corrección y su fundamento teórico.
- ✦ Describir las dificultades que implicó la construcción de un instrumento que reflejara el constructo “calidad de la escritura ensayística de los estudiantes preuniversitarios”, según el fundamento teórico del MCC.
- ✦ Proponer su utilidad para que los estudiantes desarrollen una mayor conciencia de las demandas de las tareas escritas, y para que los profesores favorezcan el desarrollo de esta habilidad mediante una retroalimentación focalizada.

En la primera parte del artículo se presentan los distintos tipos de instrumentos de puntuación (listas de comprobación, rúbricas holísticas y rúbricas analíticas). En la segunda se describe el proceso de construcción de la pauta de corrección (rúbrica) del EXHALING, así como el constructo teórico que la fundamenta. En la tercera se expone la versión final de instrumento. Finalmente se argumenta a favor de su utilidad para desarrollar la habilidad escrita en el aula preuniversitaria.

Los distintos tipos de instrumentos de puntuación

La escritura involucra diversos procesos cognitivos y motivacionales: establecer objetivos, transformar contenidos, seguir una línea temática, tener en cuenta a la audiencia, conocer las características de los distintos géneros discursivos, autorregularse y tomar múltiples decisiones para alcanzar el objetivo (Bereiter y Scardamalia, 1987; Flower y Hayes, 1981,1984; Flower, 1998; Hayes, 2000; Dempsey, 2009).

A la complejidad inherente de esta habilidad lingüística se suma el reto de evaluar objetivamente el constructo denominado “calidad de la escritura”. Desde la década de los cuarenta, la evaluación de la escritura ha transitado del interés por el texto mismo hacia el paradigma cognitivo y hacia el proceso de redacción, entremezclado con el contexto sociocultural y con teorías discursivas y dialógicas del texto (Nystrand, Greene y Wiemelt, 1993, citado por Sperling y Freedman, 2001:371). Según Galán, dos afirmaciones dan una visión sucinta del tema: 1) Hay poco consenso sobre qué es “escribir bien”; 2) En todos los casos, desde sus inicios, evaluar la redacción ha sido “un problema espinoso” (Scardamalia y Bereiter, 1986:779, citado por Galán, 2008: 75).

Puesto que se trata de procesos cognitivos, la evaluación de la escritura se lleva a cabo utilizando esquemas, guías o escalas de puntuación que establecen explíci-

tamente los criterios de actuación para clasificar y organizar observaciones respecto a diferentes niveles de habilidad, conducta y calidad del producto (Bargainnier, 2003:1). Las escalas incluyen el enunciado del criterio por evaluar y un sistema apropiado de puntuación. El primero debe corresponder a un criterio significativo y estar formulado con un lenguaje descriptivo, claro y accesible. Debe permitir discriminar los diferentes niveles de actuación de manera cualitativa, evidente y no arbitraria. Por otra parte, la puntuación es la representación explícita de la teoría subyacente (es decir, ciertos aspectos son más o menos importantes que otros); tiene consecuencias en la ponderación final y por tanto en la toma de decisiones. Las escalas bien diseñadas, además, se enfocan positivamente en los logros; diferencian entre actuación, producto y esfuerzo; y exhiben atributos psicométricos de validez universal y confiabilidad (Bargainnier, 2003:4).

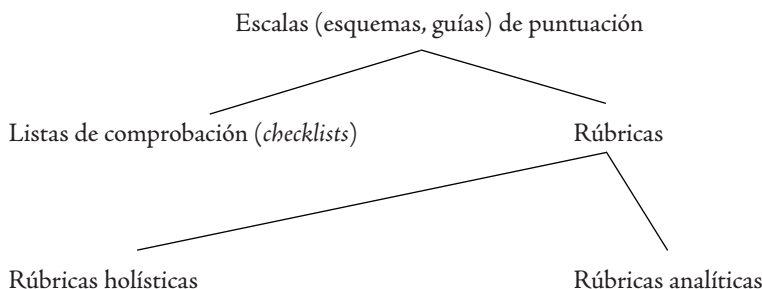
Además de describir los logros de los estudiantes en términos de un conjunto de criterios de actuación y de servir como definiciones operativas del constructo teórico, de las creencias acerca de lo que es escribir, North y Schneider (1998) enumeran las siguientes funciones que cumplen las escalas de puntuación: *a*) proveer criterios para que los estudiantes puedan confrontar sus escritos; *b*) reducir la subjetividad de las calificaciones suministrando un estándar común y significativo para los evaluadores; *c*) proporcionar guías para la construcción de exámenes; *d*) reportar resultados homogéneos, en términos de un mismo instrumento; *e*) proveer vínculos institucionales coherentes entre la evaluación diagnóstica, el plan de estudios, los materiales, la evaluación de proceso y la certificación; *f*) establecer un marco de referencia que describa los logros de un sistema educativo en términos significativos para los involucrados; *g*) posibilitar la comparación entre sistemas o poblaciones usando una medida común.

Mediante la utilización de esquemas de puntuación se pretende reducir la subjetividad inherente a la evaluación del constructo denominado “calidad de la escritura” o “escribir bien”. Las escalas o guías de puntuación pueden ser de distintos tipos. Las listas de comprobación (*checklists*) verifican la presencia o ausencia de un atributo, son unidimensionales y se utilizan preferentemente para evaluar actuaciones. Las rúbricas (*rubrics*) indican el grado en el que se alcanza un criterio, son multidimensionales y se usan para la valoración tanto de las actuaciones como de los productos resultantes de la ejecución de una tarea.

Las rúbricas, a su vez, pueden ser holísticas o analíticas. Las primeras requieren la evaluación del proceso o producto en su totalidad, sin juzgar separadamente las partes que lo componen. Generalmente se utilizan cuando los errores en alguna parte del proceso pueden tolerarse en aras de la alta calidad de la actuación completa. La

principal ventaja de las rúbricas holísticas radica en su carácter práctico, ya que “su aplicación requiere una sola lectura rápida del escrito por parte de cada uno de los evaluadores que participan en el proceso. Por otra parte, con las rúbricas analíticas primero se asignan puntajes separados a las partes del producto o actuación y luego se suman para obtener un puntaje total. La desventaja de las rúbricas analíticas es que su aplicación demanda varias lecturas del escrito por parte de cada evaluador, con lo cual el proceso se vuelve considerablemente lento. Ambos tipos de rúbricas poseen ventajas y desventajas. A continuación se presenta un cuadro sinóptico que las resume:

CUADRO I



Ventajas	Desventajas	Ventajas	Desventajas
El énfasis está en lo que los escritores hacen bien y no en sus deficiencias (Cohen, 1994).	El puntaje no ofrece información diagnóstica; se reduce la confiabilidad del instrumento (Caruso, 1996).	Las categorías no se encuentran colapsadas en un puntaje inflado; se puede entrenar fácilmente a los codificadores (Cohen 1994).	Evaluar utilizando una escala analítica puede influir a la hora de utilizar otra; este tipo de escalas no suele ser informativa para quien es evaluado (Myford y Wolfe, 2003).
La validez es mayor porque refleja una reacción auténtica y personal del lector hacia el texto (White, 1984).	El puntaje depende más del evaluador que de las propiedades del texto (Hamp-Lyons, 2003).	Es posible generalizar diversas tareas de producción escrita (Weigle, 2002).	Su desarrollo requiere mucho tiempo y puede resultar un producto costoso (Hamp-Lyons, 2003; Weigle, 2002).
El puntaje se determina rápidamente (Weigle, 2002).	La información sobre los aspectos de escritura que merecen mayor atención es insuficiente (Nelson y Van Meter, 2007).	La confiabilidad que se obtiene es mejor (Huot, 1996; Knoch, 2009).	Las subhabilidades vinculadas con la escritura no se pueden separar (White, 1984).

Ventajas	Desventajas	Ventajas	Desventajas
		Puede ayudar al escritor a identificar sus fortalezas y sus debilidades; ofrece información diagnóstica (Bacha, 2001; Carr 2000).	Los evaluadores pueden juzgar la escala de manera holística para empatar percepciones generales (Nakamura, 2004).

(Adaptado de Mertler 2001 y Becker, 2011: 115-116)

En suma, las escalas de puntuación cumplen diferentes fines. Para la evaluación del constructo subyacente a la escritura, cuando se trata de exámenes aplicados a gran escala, se utilizan fundamentalmente rúbricas holísticas; para la evaluación focalizada de las fortalezas y debilidades de los productos se opta por las rúbricas analíticas, cuyo diseño y aplicación resultan más largos y laboriosos.

Proceso de construcción de la pauta de corrección del EXHALING

El constructo del EXHALING se organiza en tres niveles de documentación: marco de referencia, marco teórico subyacente y especificaciones (Jurado, 2014). Puesto que se trataba de un examen referido a un criterio, se adoptó como tal el Marco Curricular Común (MCC) del bachillerato, publicado en 2008 por la Secretaría de Educación Pública. En particular se trabajó con la categoría “se expresa y comunica”. Los atributos de dicha categoría fueron sistematizados en una retícula a partir de la cual se definieron tablas de especificaciones para las pruebas de respuesta seleccionada (comprensión auditiva, comprensión de lectura y conciencia lingüística) y tareas para la prueba de respuesta construida (expresión escrita).

El marco teórico subyacente al marco de referencia, es decir, subyacente al MCC del bachillerato, tiene tres pilares: *a*) el concepto de competencia comunicativa; *b*) una teoría psicosociolingüística que enfoca los procesos de lectura y escritura desde una perspectiva cognitivo-discursiva, y *c*) el concepto de conciencia lingüística. Sólo los dos primeros se relacionan con la pauta de corrección.

La competencia comunicativa centra su interés en la identificación de la forma como se producen los intercambios comunicativos y lingüísticos en la vida cotidiana, la determinación de los elementos involucrados en la expresión y comprensión de los mensajes, así como en la indagación del papel que desempeña cada uno de los

elementos que intervienen en la interacción comunicativa (Canale y Swain, 1980; Canale, 1983; Bachman, 1990; Lomas, 1993; MCER, 2000). Los distintos modelos de competencia comunicativa coinciden al menos en el reconocimiento de cuatro subcompetencias: lingüística (la capacidad innata para hablar y el conocimiento gramatical de la lengua), sociolingüística (conocimiento de normas socioculturales en diferentes usos lingüísticos), pragmático-discursiva (conocimientos y habilidades para comprender y producir diferentes tipos de discursos) y estratégica (recursos, procedimientos, planes o acciones que pone en práctica el emisor de forma potencialmente consciente para maximizar la eficacia de su aprendizaje o de su comunicación).

Desde la teoría psicosociolingüística que constituye el segundo pilar del constructo, la habilidad de escribir es concebida como una compleja actividad cultural, social, cognoscitiva e interactiva que va más allá de lo motriz, ya que, por una parte, conlleva un componente dinámico que permite al escritor el manejo de la situación de producción escrita, y por otra, le posibilita —mediante la conducta estratégica— efectuar un procesamiento flexible de los recursos disponibles. Específicamente en el ámbito académico, se trata de una actividad de construcción de significados y, más importante aun, una instancia de generación de conocimientos y desarrollo del pensamiento e inteligencia del sujeto escritor (Parodi, 2003; Carlino, 2004; Cisneros, 2011).

Para producir discursos o textos con un sentido pleno, intencionado y convincente, un escritor académico experto vincula microestructuras (relaciones entre oraciones, especialmente de coherencia referencial y textual) con macroestructuras (relaciones semánticas globales del contenido del texto que reflejan el sentido general del discurso) y con superestructuras textuales (esquemas que caracterizan los tipos de texto y que conducen su interpretación y producción).

Se trata de una teoría de comprensión (no exclusivamente de producción) textual que descansa en dos supuestos básicos relacionados con el procesamiento lingüístico (Parodi, 2003):

- a. Supuesto cognoscitivo. Cuando un lector se enfrenta a un texto, construye una representación mental de su comprensión basada en sus propias categorías semánticas; por lo tanto, para que se lleve a cabo el proceso de interpretación se requiere que el sujeto posea algunos conocimientos sobre el tema en cuestión y sobre cómo realizar el proceso de lectura.
- b. Supuesto contextual. La comprensión del discurso no ocurre en el vacío, sino que el lector se enfrenta a la comprensión de un texto como parte de un evento social contextual que incluye las normas, valores, actitudes y convenciones sobre los participantes y sus respectivas interacciones como

elementos de una situación concreta de comunicación, en un entorno social determinado.

Como puede apreciarse, el segundo supuesto se conecta directamente con el concepto de competencia comunicativa subyacente a la categoría “se expresa y comunica” del MCC del bachillerato. El siguiente esquema tomado de Jurado (2014) sistematiza las relaciones entre los pilares y los supuestos del constructo subyacente con la pauta de corrección de las pruebas escritas:

	Base psicosociolingüística	Subcompetencias comunicativas	Rubros
Estructura textual	Microestructura	Lingüística	Rango y control léxico Precisión gramatical Complejidad sintáctica Ortografía Acentuación Puntuación Cohesión Estructura del párrafo
	Macroestructura	Discursiva	Título
	Superestructura		Coherencia
Modelo de procesamiento lingüístico	Supuesto cognoscitivo	Estratégica	Introducción, tesis, argumentos, contrargumentos y conclusión
	Supuesto contextual	Sociolingüística	Planeación Monitoreo
			Adecuación

Una vez definido el constructo, se procedió a hacerlo operativo tanto en el examen como en la escala de puntuación para evaluar las tareas escritas. Varias versiones fueron confeccionadas *a priori* por el equipo de especialistas (Oropeza *et al.*, 2014). La primera de ellas estaba muy apegada al constructo, pero resultaba poco práctica, entre otras razones, porque evaluaba primero la competencia lingüística y al final la

sociolingüística y la estratégica, que resultaban más rápidas de calificar. Para facilitar la aplicación práctica de la escala, se determinó invertir el orden y también eliminar, o mejor dicho “desatar”, los macrorrubros correspondientes a la competencia lingüística y a la pragmática, para dejar solo una lista de criterios. Con estas dos medidas se sacrificó la manifestación evidente del constructo, pero se ganó en “practicidad”.

El prurito de plasmar el constructo confrontó al equipo con varias contradicciones. Por una parte, por tratarse de un examen aplicado a gran escala, los 4 041 ensayos del EXHALING debieron ser evaluados con una rúbrica holística; sin embargo, se trataba también de un examen diagnóstico. Este tipo de instrumentos se caracteriza por enfocar elementos específicos más que habilidades globales. Su objetivo es identificar las fortalezas y debilidades en el conocimiento y uso de la lengua. Proporcionan un reporte de resultados que posibilita el análisis detallado de las respuestas a los ítems y a las tareas, por tanto, funcionan también como retroalimentadores (Alderson, 2005). Ante esta contradicción, el equipo de especialistas del EXHALING decidió explorar tanto el camino holístico como el analítico. En el anexo 1 aparece una muestra de los descriptores holísticos para las puntuaciones máxima y mínima.

Por otra parte, se encontró que algunos rubros o aspectos de calidad que se pretendía evaluar podían ser desglosados en niveles de puntaje, pero otros tenían que ser valorados solo por su presencia o ausencia. Esta multidimensionalidad condujo a la combinación de una escala de puntuación analítica con una serie de enunciados en forma de lista de comprobación para ciertos aspectos de calidad correspondientes a la subcompetencia discursiva.

Otra cuestión que tuvo que ser resuelta fue el hecho de que, debido a restricciones presupuestales, no fue posible contar con un equipo de evaluadores entrenados (codificadores) expertos, razón por la cual se conformó un grupo compuesto por profesores experimentados en la enseñanza de lengua, estudiantes recién egresados y prestadores de servicio social que cursaban los últimos semestres de carreras de lingüística y letras modernas. La imposibilidad de desarrollar un trabajo concentrado y sostenido condujo al desarrollo de un tipo de escala analítica, pero muy práctica. Las escalas de este tipo suelen ser detalladas y a veces con complejas explicaciones del procedimiento de puntaje. En el caso del EXHALING, se trató de confeccionar un instrumento práctico, que fue complementado por una guía explicativa.

Durante el taller de 20 horas de entrenamiento de evaluadores (codificadores), la discusión argumentada basada en la aplicación de la escala llevó la enunciación cuantificada de algunos criterios (p. ej. “hasta tres errores”; “más de tres errores”), así como a la eliminación de la valoración holística por la discrepancia de más de dos puntos que arrojaba la puntuación entre evaluadores entrenados (*raters*), quienes

se enfrentaban al dilema de valorar textos que presentaban rasgos correspondientes a niveles distintos (p. ej. el dominio de convenciones formales, pero vocabulario pobre; o mínimos errores de ortografía, pero ideas organizadas con palabras de enlace básicas). Debido a estas dificultades, se consideró que el instrumento analítico resultaba más conveniente. Las escalas fueron refinadas repetidamente para tratar de alcanzar el mayor acuerdo entre evaluadores entrenados. La versión del instrumento que se aplicó a los 4 041 ensayos del EXHALING quedó conformada por 22 criterios, rubros o aspectos de calidad (anexo 2).

Finalmente, cabe aclarar que en la publicación de los resultados del proyecto EXHALING (González, 2014), el equipo de especialistas decidió adoptar el término “pauta de corrección”, debido a que el producto final no corresponde a lo que de manera ortodoxa podría considerarse una rúbrica.

Componentes de la pauta de corrección del EXHALING

De los 22 rubros, criterios o aspectos de calidad que conforman la pauta de corrección EXHALING, 11 fueron formulados como lista de comprobación y 11 con formato de niveles de logro cuyo peso varió entre tres y dos puntos. Las escalas que fueron ponderadas con tres puntos corresponden a la competencia estratégica, la complejidad sintáctica y el uso apropiado de signos de puntuación. Las nueve restantes oscilaron en el rango de cero a dos. La valoración de la competencia discursiva se desdobló en tres niveles de logro para los rubros de coherencia, cohesión y párrafos, con un máximo de dos puntos; más nueve rubros enunciados con formato de lista de comprobación, con valores de un punto, y de dos para las operaciones discursivas.

Dado que la ponderación representa explícitamente el constructo (es decir, ciertos aspectos son más o menos importantes que otros), se determinó la siguiente distribución:

Competencia sociolingüística	2
Competencia estratégica	3
Competencia lingüística	14
Competencia discursiva	19
Total	38

Debido a la diversidad de formatos que contiene, la pauta se analizó psicométricamente con la extensión *Rasch Rating Scale Model* (RSM) de Andrich (1978). En

cuanto a la confiabilidad (consistencia en los puntajes concedidos por los evaluadores entrenados a lo largo de repetidas mediciones), el análisis mostró gran consistencia, excepto en los criterios 18 y 19, correspondientes a la evaluación de operaciones discursivas, razón por la cual fueron eliminados del reporte final de resultados, en el cual se dio cuenta solo de 20 aspectos de calidad en los ensayos de escritura. Por lo que respecta al análisis de validez (grado en el cual las puntuaciones reflejan el constructo), la media mostró gran consistencia con el resto de las pruebas del EXHALING, excepto por un desajuste de 0.05 unidades en la rúbrica 8 (uso de signos de puntuación). El siguiente mapa de Wright tomado de Montesinos (2014) muestra las medias y los niveles de dificultad de los 20 criterios en relación con la habilidad de los sustentantes:

MAPA I

TABLE 12.2 CRAM RATING FINAL Ensayos Escritura ZOU760WS.TXT Oct 1 11:12 2013
INPUT: 4041 RATING 22 ASPECTO REPORTED: 4041 RATING 20 ASPECTO 13 CATS
WINSTEPS 3.80.1

MEASURE RATING - MAP - ASPECTO

<more>|<rare>

80 +

|

|

.|

.|

.|

.|

.|

70 . +

.|T

.# | 14 SUFI

.# T|

.|

.## | 16 CONT

.## |

60 .## + 09 TITU

.##### |S 02 COM3 17 CONC

.##### S| 07 ACEN

.##### | 08 PUNT

|

.##### | 01 COM2

.|

(51.63)

.##### |M 03 RANG

 50 .##### + 06 ORTO 10 INTR 20 COH1

```

##### M| 12 ARGU(49.57)
##### |
##### | 15 VERO 22 COH3
##### | 04 PREC 13 PERT
|S 11 TESI 21 COH2
##### |
##### S|
40 . +
##### |
#### |
|
### |T
| 05 COMP
## T|
|
30 .# +
.|
|
.|
|
20 . +
|
|
.|
|
10 +
.|
|
0 +
<less>|<frequent>
EACH "#" IS 26; EACH "" IS 1 TO 25

```

Como se aprecia en el mapa, el posicionamiento de los sustentantes arrojó una media de 49.57, cercana a la media de exigencia de pauta, que fue de 51.63. La rúbrica con menor índice de dificultad fue la de complejidad sintáctica, y la de mayor corresponde a la de suficiencia de argumentos. En el rango medio se ubicaron las correspondientes al control léxico, ortografía, introducción y coherencia.

Utilidad de la pauta de corrección EXHALING para desarrollar la habilidad escrita en el aula preuniversitaria

Como ya se mencionó en la primera sección de este artículo, las escalas analíticas tienen como una de sus principales características la de servir para retroalimentar

procesos. Varios autores coinciden en que se trata de un recurso poderoso para desarrollar en los estudiantes la autoconciencia de la tarea y por tanto mejorar su desempeño escrito:

Three reviews of the research on the effectiveness of formative assessment to increase student achievement have consistently demonstrated that assessment providing students with feedback about their performances increases subsequent performances (Natriello, 1987; Crooks, 1988; Black & Wiliam, 1998). A key factor in the success of formative assessment is the benefit to student achievement resulting from self-assessment based on an understanding of the goals of learning and the criteria by which students are evaluated (Fontana & Fernandes, 1994, en Kozlow & Bellamy, 2004:5)

Asimismo, los beneficios son también para los profesores, quienes pueden contar con un instrumento para medir el grado en el que un criterio específico se ha alcanzado, y por tanto para retroalimentar de manera focalizada y fundamentada los trabajos de sus estudiantes. Les sirve además para recabar información sistematizada y de ese modo llevar a cabo investigaciones (incluso se podría pensar a futuro en una investigación para desarrollar una aplicación tecnológica accesible para evaluar ensayos):

Consistent use of well-designed rubrics significantly improves the facilitation of learning by providing both students and instructors with clarity and commonality of purpose. Students can better validate their own progress, and instructors can fairly and consistently document the students' skills and growth. Using rubrics across programs and/or course sequences can also insure consistent measurement of quality of performance by students who have different instructors. Rubrics can provide a well-founded measurement system for improving teacher performance and collecting data for research projects. Creating and using quality rubrics is well worth the considerable time and effort involved (Bargainier, 2003:4).

La pauta de corrección del EXHALING no tuvo en sus orígenes esa finalidad; no obstante, las necesidades derivadas de la investigación empírica condujeron a la confección de un instrumento ecléctico y relativamente amigable que podría ser utilizado —con algunos ajustes— para desarrollar en los estudiantes un conocimiento metacognitivo y metalingüístico útil para guiar la escritura de ensayos argumentativos, una de las tareas de escritura más comunes en el nivel bachillerato,

que les proporciona la base para elaborar artículos académicos y reportar sus investigaciones en el nivel universitario.

Conclusiones

Las escalas de puntuación son instrumentos ampliamente utilizados en el campo de la evaluación tanto de procesos como de productos. Existe una multitud de modelos validados con alto rigor metodológico y ampliamente utilizados en distintos contextos académicos para medir el constructo “calidad de la escritura”.

El EXHALING fue un examen diagnóstico de las habilidades lingüísticas de los jóvenes que ingresan al nivel universitario. La pauta de corrección de su prueba escrita fue confeccionada atendiendo a la necesidad de evaluar una gran cantidad de ensayos, recabando información que respondiera a un constructo teórico específico (el MCC del bachillerato). Como todo instrumento de calidad, su construcción y aplicación requirió el esfuerzo de varios especialistas.

Aunque estamos conscientes de la dificultad para transferir o adaptar un instrumento de evaluación de producto a un contexto de instrucción, también tenemos la certeza de que, tanto por su metodología como por sus peculiaridades en cuanto a eclecticismo y practicidad, la pauta de corrección EXHALING puede ser utilizada, con algunos ajustes, como instrumento relativamente consensuado de retroalimentación para distinguir niveles de habilidad, mejorar los escritos de los jóvenes bachilleres, fundamentar la retroalimentación de los docentes y acopiar información sistemática para investigaciones.

El trabajo de afinación de un instrumento de esta naturaleza es inagotable. Una nueva versión (que constituiría una base bastante sólida para el diseño de una herramienta tecnológica) tendría que tomar en cuenta los siguientes ajustes: en primer lugar, la fusión de la pauta de corrección con algunos extractos de la guía explicativa, para que los usuarios —alumnos y maestros— contaran con un solo y amigable instrumento. En segundo término, es necesario complementar los niveles de logro para valorar con mayor precisión el contenido y el aparato crítico (citas y referencias bibliográficas). Finalmente, se necesita buscar la manera adecuada para medir los criterios cuyos índices resultaron débiles y excluidos por el análisis psicométrico (complejidad sintáctica, puntuación y operaciones discursivas).

Bibliografía

- Bachman, L. (1990), *Fundamental considerations in language testing*, Oxford University Press, Oxford.
- Bachman, L. y A. Palmer, A. (1996), *Language testing in practice*, Oxford University Press, Oxford.
- Bargainnier, S. (2003), Fundamentals of rubrics. *Pacific Crest* (pp. 1-4). [versión digital], <https://www.webpages.uidaho.edu/ele/scholars/practices/Evaluating_Projects/Resources/Using_Rubrics.pdf> [22 de marzo de 2014].
- Canale, M. (1983), On some dimensions of language proficiency, en J. W. J. Oller (Ed.), *Issues in language testing research* (pp. 333-342), Newbury House, Rowley, MA.
- Canale, M. y M. Swain (1980), "Theoretical bases of communicative approaches to second language teaching and testing", *Applied Linguistics*, 1, pp. 1-47.
- Flower, L. (1990), *Reading-to-write: exploring a cognitive and social process*, Oxford, University Press, Oxford.
- Galán, R. (2008), *Factores en el dominio de la redacción al ingreso en la universidad*, tesis de doctorado inédita, Universidad Anáhuac, México D. F.
- Hamp-Lyons, L. (1990), "Second language writing: Assessment issues", en B. Kroll (ed.), *Second language writing: Research insights for the classroom*, Cambridge University Press, Nueva York, pp. 69-87.
- (1991), "Scoring procedures for ESL contexts", en L. Hamp-Lyons (ed.), *Assessing second language writing in academic contexts*, Ablex, Norwood, NJ, pp. 241-276.
- Jurado, M. (2014), "El examen de habilidades lingüísticas (EXHALING)", en Rosa O. González (coord.), *Habilidades lingüísticas de los estudiantes de primer ingreso a las instituciones de educación superior del área metropolitana de la Ciudad de México*, ANUIES, México.
- (2010a), *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*, <<http://cvc.cervantes.es/obref/marco/>>, Centro Virtual Cervantes. Marco de referencia europeo [9 de enero de 2010].
- (2010b), *Marco Curricular Común (MCC) para la Educación Media Superior*, elaborado por la SEP y decretado en el *Diario Oficial de la Federación*, 23 de septiembre de 2008. <http://www.oei.es/pdfs/reforma_educacion_media_mexico.pdf> [15 de abril de 2010].

- Kozlow, M. y P. Bellamy (2004), *Experimental study on the impact of the “6 +1 Trait” Writing Model on student achievement in writing*, Northwest Regional Educational Laboratory, Portland, OR.
- Messick, S. (1996), “Validity and washback in language testing”, *Language Testing*, 13(3), pp. 241-256.
- Oropeza, V.; M. Galán, M. Jurado y V. Vázquez (2014), “La expresión escrita y su pauta de corrección”, en Rosa O. González (coord.), *Habilidades lingüísticas de los estudiantes de primer ingreso a las instituciones de educación superior del área metropolitana de la Ciudad de México*, ANUIES, México.
- Nitko, A. J. (2001), *Educational assessment of students*, 3ª ed., Merrill, Upper Saddle River, NJ.
- North, B. y Schneider, G. (1998), “Scaling descriptors for language proficiency scales”, *Language Testing*, 15(2), pp. 217-263.
- Shohamy, E., C. Gordon y R. Kraemer (1992), “The effect of rater background and training on the reliability of direct writing tests”, *The Modern Language Journal*, 76(1), pp. 28-33.
- UNESCO (1983), *Guía para la redacción de artículos científicos destinados a la publicación*, 2ª. ed., París [versión digital] <www5.uva.es/iuu/DownLoads/06_Guia_UNESCO.pdf>. [13 de enero de 2014].
- Weigle, S. C. (2002), *Assessing writing*, Cambridge University Press, Cambridge.

ANEXO I
MUESTRA DE DOS NIVELES DE PUNTAJE DE LA ESCALA HOLÍSTICA

<p>El texto es sobresaliente porque presenta</p> <ul style="list-style-type: none"> • Contenidos principales y pertinentes. • Extensión adecuada para cada tipo de contenido. • Dominio de convenciones formales. • Claro hilo conductor entre las ideas. • Enlaces bien empleados y variados. • Vocabulario variado y preciso. • Precisión gramatical con errores mínimos. • Mínimos errores de ortografía y puntuación. 	4
---	---

[...]

<p>El texto es malo porque presenta</p> <ul style="list-style-type: none"> • Omisión o tratamiento incorrecto de contenidos importantes. • Información no pertinente. • Ausencia de convenciones formales. • Registro inadecuado. • Vocabulario limitado e insuficiente. • Múltiples errores gramaticales. • Múltiples errores de ortografía y puntuación. 	0
---	---

ANEXO 2
 TALLER PARA LA EVALUACIÓN DE EXPRESIÓN ESCRITA
 UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
 CENTRO DE ENSEÑANZA PARA EXTRANJEROS
 12-16 DE MARZO DE 2012¹

(1) Competencia sociolingüística	Puntos
Adecuado sin fallas	2
Adecuado hasta con dos fallas	1
Inadecuado (más de dos fallas)	0
(2) Competencia estratégica	
Planeación y monitoreo	3
Planeación	2
Monitoreo	1
Ninguno	0
(3) Rango y control léxico	Puntos
Vocabulario adecuado, variado y especializado. Menos de tres errores.	2
Vocabulario adecuado, pero con repeticiones e imprecisiones. Tres errores.	1
Vocabulario pobre y repetitivo. Más de tres errores.	0

¹ La rúbrica de base fue propuesta por el equipo de especialistas en español conformado por Margarita Galán, Martha Jurado, Irma Munguía, Viviana Oropeza, Elia Sánchez y Violeta Vázquez. En el taller de entrenamiento, coordinado por Georgina Barraza, se ajustó y complementó en colaboración con los asistentes.

(4) Precisión gramatical	Puntos
Sin errores (verbo, adverbio, preposición, gerundio, adjetivo, locución prepositiva y locución conjuntiva)	2
Hasta tres errores	1
Más de tres errores	0
(5) Complejidad sintáctica	Puntos
Uso de siete o más nexos subordinantes distintos, sin errores (mal empleo de nexos o problemas de correlación entre ellos).	3
Uso de seis nexos subordinantes distintos, sin error (o siete nexos subordinantes distintos hasta con un error).	2
Uso de seis o menos nexos subordinantes distintos con o sin errores.	1
(6) Ortografía	Puntos
Sin errores	2
Hasta tres errores (en palabras distintas)	1
Más de tres errores (en palabras distintas)	0
(7) Acentuación	Puntos
Sin errores	2
Hasta tres errores (en palabras distintas)	1
Más de tres errores (en palabras distintas)	0
(8) Puntuación (ponderación acumulativa)	Puntos
Grupos 1 (, .) y 3 ; : () — ... "" sin errores*	3
Grupo 1 sin errores y grupo 3 con error(es)	2
Grupo 1 sin errores.	1
Grupo 1 con errores	0

NOTA 1: El grupo 2 (signos de admiración e interrogación) se integran al Grupo 1, aunque pueden o no aparecer en el texto.

NOTA 2: Se califica el mal empleo de signos de puntuación. NO SU AUSENCIA. La única ausencia que se califica como error en esta rúbrica es la falta de coma en enumeraciones.

*La ausencia de signos del grupo 3 se marca a favor del alumno.

	Sí = 1	No = 0
Desarrollo de ideas y precisión proposicional		
(9) ¿Hay un título que presente el tema a tratar?		
(10) ¿Hay una introducción que presente el marco temático?		
(11) ¿Retoma la tesis propuesta?		
(12) ¿Presenta, al menos, dos argumentos?		
(13) ¿Los argumentos son pertinentes?		
(14) ¿Los argumentos son suficientes?		
(15) ¿Los argumentos son verosímiles?		
(16) ¿Presenta al menos un contra argumento?		
(17) ¿Hay una conclusión que sintetiza el planteamiento o reelabora la tesis? (también se obtiene cero si el texto tiene cierre, pero no tiene conclusión).		
	Sí = 2	No = 0
(18) Operaciones de primer tipo (secuencia progresiva, ejemplificación, definición, clasificación, comparación y contraste).		
(19) Operaciones de segundo tipo (analogía compleja, cita de autoridades, cita de datos objetivos, pregunta o figura retórica, ironía, golpe de efecto en las conclusiones).		
(20) Coherencia	Puntos	
Hilo conductor claro y fluido	2	
Enlace elemental entre párrafos	1	
Sin enlace o enlace incoherente	0	
(21) Cohesión	Puntos	
Dos errores (deícticos, pronombres, elipsis, concordancia, conjunciones a nivel de frase, no de oración)	2	
Tres errores	1	
Más de tres errores	0	

NOTA: En este apartado se califica también como error la ausencia flagrante de signos de puntuación que afectan la comprensión del párrafo.

(22) Organización en párrafos	Puntos
El texto presenta más de 3 párrafos (una idea principal por cada párrafo). Se permite un error	2
El texto presenta tres párrafos (o más de tres párrafos, con dos errores)	1
El texto presenta menos de tres párrafos, o tres párrafos con un error, o más de tres párrafos con más de dos errores	0