

Validez y Justicia: hacia una evaluación significativa en pruebas estandarizadas de escritura

▮ Federico Navarro *

▮ Natalia Ávila Reyes **

▮ Gabriela Gómez Vera ***

Resumen

Hasta hace poco, la validez de las pruebas estandarizadas de escritura se basaba en la confiabilidad inter-correctores, las tareas no estaban informadas por teoría e investigación, los textos eran evaluados mediante indicadores indirectos o rúbricas analíticas que fragmentaban sus componentes y, sobre todo, la validación contextual, la justicia y las consecuencias del uso de resultados no eran consideradas. Este artículo presenta el desarrollo de una evaluación de escritura para proveer información del desempeño escrito al inicio de la educación superior en Chile. Incluye tareas para escribir con fines académicos y de participación ciudadana; detalla la situación, los propósitos, las audiencias y las expectativas; habilita la planificación y revisión; y evalúa los textos mediante una rúbrica holística centrada en el cumplimiento de objetivos. Los resultados de su aplicación muestran un alto nivel de interés de los estudiantes y una distribución normal de los desempeños.

Palabras-clave: Escritura. Evaluación estandarizada. Educación superior. Géneros discursivos. Validez.

* Doctor en Lingüística, Universidad de Valladolid (España); profesor de la Universidad de O'Higgins (Chile). E-mail: navarro@uoh.cl.

** Doctora en Educación, University of California Santa Barbara (Estados Unidos); profesora de la Pontificia Universidad Católica de Chile (Chile); E-mail: naavila@uc.cl.

*** Doctora en Educación, Université de Bourgogne (Francia); profesora de la Universidad de O'Higgins (Chile); E-mail: gabriela.gomez@uoh.cl.

Se agradece el financiamiento otorgado por el Proyecto FONDEF ID16I10090, por el Proyecto FONDECYT regular N° 1191069, por el PMI UCH1501 y por el Fondo Basal para Centros de Excelencia proyecto FB0003 de PIA-CONICYT. Asimismo, se agradece el apoyo del equipo directivo y del Comité de Lenguaje del Departamento de Evaluación, Medición y Registro Educacional de la Universidad de Chile, en particular de su directora Leonor Varas y de Karen Fernández.

1. Introducción

La escritura ha sido conceptualizada como un artefacto o producto, como una competencia individual para actuar, como una actividad situada en contextos sociales y culturales, como una tecnología con una larga historia de desarrollo o como un modo de organización social (PRIOR; LUNSFORD, 2008). Estas diferentes miradas sobre la escritura coinciden en que ocupan un lugar central en la sociedad de la información, de la comunicación y del conocimiento (STARKE-MEYERRING; PARÉ, 2011); que se caracteriza por ser compleja, diversa, situada y retórica; que opera en las dimensiones sociales, cognitivas e identitarias (ADLER-KASSNER; WARDLE, 2015); y, hoy en día, un escritor competente debe poder comprender, aprender, negociar y adecuarse a las sofisticadas expectativas comunicativas de entornos distintos (BAZERMAN, 2013).

En las últimas décadas se ha incrementado el uso de instrumentos estandarizados de medición de la competencia de escritura en estudiantes de diferentes niveles educativos con fines diversos, tanto de diagnóstico del estudiante, de la institución o del sistema educativo en general, como para fines de selectividad y asignación de nivel (*placement*) (PERELMAN, 2018). En Latinoamérica, sin embargo, las pruebas de escritura siguen siendo mucho menos frecuentes que las pruebas de lectura, y en general se aplican en niveles educativos escolares. La expansión del sistema de educación superior en las últimas décadas (CHIROLEU; MARQUINA, 2017), junto con el interés social e institucional tanto en promover instrumentos de acreditación y aseguramiento de la calidad de carreras y programas (LEMAITRE, 2015) como en fomentar el apoyo y acompañamiento de los aprendizajes estudiantiles (SANTELICES; GALLEGUILLOS; CATALÁN AVENDAÑO, 2015), han impulsado la creación de nuevos instrumentos de evaluación estandarizada de escritores jóvenes adultos en la finalización de la educación media y en diferentes etapas de la educación superior. Pueden mencionarse, por ejemplo, la prueba Saber Pro en Colombia o la Prueba de Selección Universitaria (PSU) en Chile, junto con otras iniciativas más bien locales de evaluación escrita aplicadas frecuentemente en las universidades chilenas (SANTELICES; GALLEGUILLOS; CATALÁN AVENDAÑO, 2015). Este tipo de pruebas estandarizadas de gran escala tienen un amplio impacto social, cultural, institucional y educativo, ya que influyen en las trayectorias formativas de los estudiantes, en el diseño curricular de carreras e instituciones y en las prácticas pedagógicas de los profesores (O'NEIL, 2012).

Sin embargo, es aún escasa la discusión teórica en la región respecto de la validez de los instrumentos de medición de la escritura en etapas avanzadas de la educación formal, en términos tanto de los fundamentos teóricos que informan los instrumentos de medición (BROAD, 2003; MISLEVY, 2018), como de la justicia social y los usos pedagógicos de los resultados de la evaluación (HUOT, 2009; PERELMAN, 2018; POE; INOUE; ELLIOT, 2018), propios de una noción contemporánea de validez (MURPHY; YANCEY, 2008; O'NEILL; MOORE; HUOT, 2009) y que han ganado centralidad a nivel internacional en el campo de la evaluación de competencias.

Este artículo tiene como propósito describir el desarrollo de una propuesta válida, situada y significativa de evaluación estandarizada de la escritura en educación superior, en torno a la elaboración de seis principios teóricos para la validez de evaluaciones estandarizadas de escritura. Primero, se ofrece un relevo teórico sistemático del conocimiento pertinente para fines evaluativos consensuado por los estudios de la escritura. Luego, se vinculan esos principios o conceptos clave con el estado del saber actual sobre pruebas de escritura con respecto a su validez, entendido tanto desde el punto de vista de su fundamentación teórica como desde su uso e impacto social, además del criterio clásico de confiabilidad y de los aspectos de tipo logístico y metodológico. Posteriormente, se discuten de forma crítica instrumentos de medición de la escritura utilizados a nivel internacional y regional que implementan algunos de estos principios. Finalmente, se ofrece una propuesta original de medición estandarizada de la competencia de escritura en el ingreso a la educación superior que ha sido diseñada por los autores desde el año 2016, siguiendo los criterios previos y en colaboración con el Departamento de Evaluación, Medición y Registro Educacional de la Universidad de Chile, en el marco de un desarrollo más amplio de un conjunto de instrumentos para evaluar competencias fundamentales y habilidades transversales necesarias para cursar exitosamente un programa de estudios en el sistema de educación superior.

2. Escritura en la sociedad del conocimiento

La escritura es una tecnología **epistémica** y **semiótica**, es decir, que construye, registra y comunica significados. Su aparición significó un avance tecnológico sustancial en la historia de la humanidad (SCHMANDT-BESSERAT; ERARD, 2008), ya que ofreció por primera vez un soporte duradero, autónomo, atesorable, transportable y decodificable por un lector ubicado en un espacio y tiempo por lo general diferentes a los del escritor.

Hoy en día, la escritura es **ubicua** en las sociedades modernas: interviene en la mayoría de las actividades, tareas, procedimientos y acontecimientos significativos (OLSON, 1998). La escritura constituye un componente vital de la infraestructura epistemológica de las organizaciones y sociedades con uso intensivo del saber (STARKE-MEYERRING; PARÉ, 2011). Por eso, la participación en la sociedad de la información, la comunicación y el conocimiento requiere cada vez más escritores avanzados y en permanente aprendizaje, que dominen escrituras distintas para propósitos diversos, y que puedan comprender, aprender, negociar y adecuarse a las expectativas de escritura de entornos nuevos (BAZERMAN, 2013).

Esta ubicuidad de la escritura es particularmente cierta en los entornos de educación formal (GARDNER; NESI, 2013): la escritura permite que los estudiantes construyan conocimiento, organizando sus ideas y vinculando lo que aprendieron con otros contenidos y experiencias (función epistémica); demuestren sus logros de aprendizaje, atravesando con éxito el control de acceso que ejercen los profesores e instituciones (función habilitante); y se comuniquen según las expectativas de las comunidades disciplinares, ofreciendo evidencias de su enculturación disciplinar (función retórica).

Pero la escritura no solo permite aprender y comunicarse en entornos escolares, sino también expresar y elaborar identidades, crear soluciones novedosas para problemas y desafíos, y asumir posiciones críticas sobre la organización del entorno, aspectos que también forman parte de la formación integral en educación superior (ver funciones de la escritura en NAVARRO, 2018). En este sentido, la escritura constituye una oportunidad para fomentar la agencia, entendida como "la capacidad socioculturalmente mediada que tienen los individuos para actuar y elegir en el marco de los efectos de las fuerzas ideológicas que han construido su subjetividad" (ZAVALA, 2011, p. 52), y el entusiasmo y compromiso de los escritores en formación (EODICE; GELLER; LERNER, 2016).

En términos más amplios, la escritura involucra varias de las llamadas competencias del s. XXI (PELLEGRINO; HILTON, 2012), fundamentales para el aprendizaje y la participación ciudadana, la discusión política y la toma de decisiones en la esfera pública (STARKE-MEYERRING; PARÉ, 2011). Por ejemplo, la escritura forma parte de competencias como el pensamiento crítico, la comunicación efectiva, la argumentación, el uso y la evaluación de información y fuentes, la resolución de problemas, la colaboración, el aprender a aprender, la autogestión y la creatividad.

Por otro lado, la escritura es **situada**: las tareas y actividades que la incluyen son cultural y situacionalmente específicas, dinámicas y cambiantes (BARTON; HAMILTON, 2000). En consecuencia, la “buena escritura” depende de prácticas, experiencias, expectativas y situaciones social, cultural e históricamente condicionadas. En palabras de Applebee, “what counts as ‘good writing’ is itself socially constructed and context-specific.” (APPLEBEE, 2007, p. 83).

La escritura es, a su vez, **compleja**: integra dimensiones múltiples y simultáneas, que requieren conocimientos, habilidades y estrategias distintas, vinculadas a factores culturales, personales, situacionales, lingüísticos y cognitivos (CAMP, 2009). Al escribir, se trabaja con aspectos mecánicos, normativos, léxicos, gramaticales, semánticos, discursivos, multimodales, tecnológicos, situacionales, retóricos, sociohistóricos y psicológicos. La escritura materializa complejos procesos de reflexión, vínculos con lecturas previas, planificación, organización en una estructura temática coherente, establecimiento de lazos cohesivos entre partes del texto, construcción de autoría y autoridad, revisión y negociación con otros, etc. (NAVARRO, 2018).

Por otro lado, la escritura no se construye en el vacío ni se compone solamente de letras, palabras y oraciones. La escritura se materializa en conjuntos de clases tipificadas de textos o **géneros discursivos** (BAJTÍN, 2005), que cumplen propósitos comunicativos en los entornos sociales, y sobre los cuales los lectores tienen expectativas y experiencias previas. Los géneros discursivos se definen como “clases recurrentes y tipificadas de textos, más o menos estables, pertenecientes a distintas esferas sociales, que sufren cambios sociales e históricos, que son reconocidos por las comunidades de uso, y que se caracterizan por los objetivos comunicativos que persiguen, vinculados a las necesidades de esas esferas sociales, y por los temas que abordan, su estructuración y desarrollo temático, y su manifestación multimodal, discursiva, gramatical y léxica” (NAVARRO, 2019, p. 15). Entender la escritura en términos de géneros discursivos permite dar cuenta de su carácter situado, complejo y significativo.

En suma, una concepción actualizada de la escritura como tecnología epistémica (construcción de conocimiento) y semiótica (transmisión de conocimiento) enfatiza su carácter complejo, significativo, diverso, tipificado, situado y transversal a los sistemas de educación formal, y estos aspectos deben informar el diseño de instrumentos de evaluación.

3. Evaluación de la escritura

La ubicuidad de la escritura en la sociedad del conocimiento, junto con su rol central en los procesos de formación escolar y en la construcción y expresión de identidades y perspectivas, hace imprescindible su incorporación a los instrumentos de medición de las competencias necesarias en programas de estudios en el sistema de educación superior. Esta necesidad contrasta con su importancia en el currículum. La escritura no suele ser enseñada como práctica situada y significativa ni aprovechada para potenciar aprendizajes en la educación básica y media (GÓMEZ VERA et al., 2016; MATEOS; MARTÍN; VILLALÓN, 2006; ROSLI; CARLINO; RONI, 2015). Por el contrario, la escritura muchas veces forma parte del currículum oculto de las instituciones educativas (SCHLEPPEGRELL, 2004), es decir, no suele explicitarse, negociarse, enseñarse ni retroalimentarse a través del currículum, sino solo evaluarse, por medio de las respuestas a pruebas, reforzando las inequidades entre los estudiantes (ROSE; MARTIN, 2012). La falta de enseñanza y ejercitación de la escritura en el aula se ve muchas veces reforzada por la influencia de pruebas estandarizadas de gran escala y con consecuencias en las trayectorias escolares que, o bien no incluyen la escritura como competencia relevante, o bien proponen tareas de escritura poco situadas o significativas (MURPHY; YANCEY, 2008; O'NEILL, 2012; PERELMAN, 2018).

Las pruebas estandarizadas de logro de escritura y de otras competencias se han convertido en un atributo propio de los sistemas educativos de los países desarrollados (PERELMAN, 2018). Sin embargo, resulta un desafío altamente complejo conjugar la necesidad de contar con instrumentos estandarizados para la evaluación de la escritura y la necesidad de responder a las concepciones actualizadas de la escritura. En este sentido, hay principios que el diseño de la evaluación debe resguardar de modo de dar cuenta del constructo que busca representar.

En primer lugar, las pruebas estandarizadas de escritura deben **evaluar de manera directa muestras reales de escritura** de estudiantes, ya que solo esta metodología puede dar cuenta de la variedad de habilidades, conocimientos y estrategias necesarias para comunicarse por escrito (CAMP, 2009; MURPHY; YANCEY, 2008). Esto permite que la evaluación pueda, en cierta medida, ser considerada evaluación **auténtica** “when we directly examine students performance on worthy intellectual tasks” (WIGGINS, 1990, p. 1), si bien no es ecológica ni incluye todos los aspectos y etapas de la escritura como

proceso en contexto. La evaluación directa de muestras reales se opone al uso de indicadores indirectos (por *proxy*), simples (y simplistas), descontextualizados, estáticos, en ocasiones arbitrarios, con foco en la aplicación o el reconocimiento de aprendizajes (WIGGINS, 1990), como sucede con la evaluación del grado de conocimiento gramatical o con el uso de respuestas de selección múltiple sobre escritura (BROAD, 2003; O'NEIL; MOORE; HUOT, 2009; WHITE, 1984).

En segundo lugar, las pruebas estandarizadas de escritura deben **solicitar tareas de escritura situadas y significativas**. Para dar cuenta de la escritura como competencia retórica compleja que cumple propósitos en contexto no pueden proponerse constructos que sostengan una concepción de la escritura genérica, simple y transversal, desvinculada de situaciones, interlocutores y fines específicos, o que no se relaciona con clases tipificadas y situadas de textos. Este principio, común a todos los niveles educativos, es especialmente cierto para los escritores jóvenes adultos y adultos que elaboran textos complejos, con estructuras sofisticadas y contenidos específicos o con cierto grado de lenguaje técnico, relacionados con la participación en actividades propias de la educación superior.

Por lo tanto, la medición estandarizada de la competencia de escritura debe relacionarse con tareas situadas, con propósitos comunicativos y sociales específicos cuyo cumplimiento se puede evaluar, destinados a lectores particulares en ciertas situaciones comunicativas, con rasgos discursivos y lingüísticos esperables para las clases tipificadas de textos o géneros discursivos de los que se trate. En términos de Camp, se trata de “simulations of real-life problems that allow for some of the open-endedness characteristic of responses to real-life situations.” (CAMP, 2009, p. 119). En el área del lenguaje, las evaluaciones fundadas en el enfoque por competencias se basan en la realización de tareas de simulación de actividades comunicativas, en un contexto social determinado, cuyo foco sea abordar y resolver un problema comunicativo (READ, 2000). Como explica Huot, “rather than base assessment decisions on the abstract and inaccurate notion of writing quality as a fixed entity, a notion which is driven by a positivist view of reality, we should define each evaluative situation and judge students upon their ability to accomplish a specific communicative task.” (HUOT, 2009, p. 169). En este sentido, los textos solicitados, si bien se producen en una situación evaluativa artificial, diferente de los contextos genuinos y ecológicos de uso, podrían funcionar y cumplir

propósitos en el mundo real y en interacción con otros, y por eso resultan significativos y pertinentes para los escritores.

En la misma línea, es recomendable proponer más de una tarea de escritura por examen, de forma tal de contar con muestras correspondientes a situaciones, propósitos y géneros discursivos diversos (CAMP, 2009). Por último, un constructo teóricamente válido no puede operar con la teoría implícita de que existe una sola textualización posible para una situación o propósito dados, no solo porque contradice el conocimiento consensuado en los estudios de la escritura, sino porque fomenta la escritura reproductiva, basada en fórmulas y entrenada para los meros fines de la evaluación. Por ejemplo, es incorrecto suponer (y evaluar en consecuencia) que un “ensayo” debe tener una estructura fija de cinco párrafos o presentar cierto número predefinido de argumentos para resultar adecuado y exitoso en todo contexto y para cualquier lector.

En tercer lugar, las pruebas estandarizadas de escritura deben colocar **al estudiante en una posición de autoridad, autonomía y agencia** sobre la temática abordada, en la que pueda aprovechar sus conocimientos, habilidades y experiencias previas para resolver la tarea (CAMP, 2009), en un diseño que promueva la “autoría propia.” (BROAD, 2003). Esta posición está informada por las últimas décadas de desarrollo de la didáctica de la escritura y su foco en “students’ individual cognitive energies and their socially positioned identities as members of culturally bound groups.” (HUOT, 2009, p. 170). En este sentido, se debe fomentar el potencial epistémico y significativo de la escritura en la construcción de mensajes y conocimientos, así como también la creatividad y expresividad de los escritores para llevar a cabo propósitos y adecuarse a contextos.

Para lograr este propósito, es preciso solicitar géneros discursivos que sean significativos y pertinentes para cualquier estudiante, que aborden temáticas específicas con las que cualquier estudiante tenga familiaridad (o bien que ofrezcan la posibilidad de seleccionar temáticas de un amplio espectro de opciones), y que valoren la diversidad de perspectivas y experiencias. A su vez, este tipo de tareas fomenta que el escritor se preocupe, comprometa, entusiasme, involucre y vincule con la tarea de escritura solicitada. Por último, este tipo de tareas y contenidos permite que todos los estudiantes tengan las mismas oportunidades de resolver la tarea.

En cuarto lugar, las pruebas estandarizadas de escritura deben **medir dimensiones escritas de alta complejidad** (PERELMAN, 2018), más allá de las dimensiones más superficiales y paradójicamente más sencillas de evaluar (LERNER, 2001), vinculadas a la

ortografía o la puntuación. Estas dimensiones, aunque importantes, no son suficientes para determinar si un texto resulta exitoso en su propósito o adecuado en su contexto. Es preciso reconocer que los textos operan en la realidad de forma integrada como artefactos complejos que cumplen propósitos y se adecuan a contextos y expectativas como un todo, y que existen dimensiones de escritura, como el uso de evidencias para construir argumentos sofisticados, que deben ser priorizadas en la evaluación, en particular cuando se trata de escritores jóvenes y adultos (APPLEBEE, 2007). Como lo explica Mislevy, ponderando la importancia de la evaluación de géneros discursivos frente a formas lingüísticas descontextualizadas, “becoming proficient means not simply knowing structures, but understanding their role in practices, using them to create and to apprehend information in the ways people communicate in those practices” (MISLEVY, 2018, p. 33). Por el contrario, el foco en dimensiones aisladas y superficiales de escritura de baja complejidad no permite dar cuenta de una concepción informada de la escritura y puede tener consecuencias negativas en prácticas de enseñanza y aprendizaje (APPLEBEE, 2007; CAMP, 2009).

De forma complementaria, las muestras directas de escritura a evaluar deben considerarse como borradores preliminares completos (APPLEBEE, 2007), que aún requieren revisiones y reescrituras sucesivas para corregirse y asentarse en su totalidad. Las condiciones de una prueba estandarizada tornan injusto esperar la producción de textos planificados y revisados en profundidad, con consulta de fuentes, en negociación y consulta con profesores y pares, como ocurre en la realidad de la escritura cotidiana. En una evaluación estandarizada, los escritores evaluados deben producir textos en un contexto que es distinto al de la producción genuina, con restricciones para planificar y revisar, o para reconsiderar y discutir con otros sus decisiones de escritura, con limitaciones de tiempo y con presiones por el nerviosismo propio de la situación de evaluación (CAMP, 2009). Si bien es cierto que los exámenes presenciales y con tiempos limitados promueven la seguridad de la evaluación (*test security*), el constructo utilizado no podrá evaluar otras dimensiones como la planificación y la revisión (PERELMAN, 2018). Además, la carga cognitiva y emotiva de los escritores estará centrada en ciertos procesos y aspectos complejos, como la estructura textual y la consideración de los propósitos y lectores, más que en aspectos de formato, prolijidad y normativa, que podrían revisarse en etapas posteriores.

Estos factores deben ser considerados al momento de la evaluación y de la ponderación de los aspectos incluidos en la rúbrica, para no penalizar de manera excesiva aspectos de ortografía y normativa o relacionados con la profundidad de los contenidos elaborados. Cabe destacar que algunas investigaciones muestran que disponer de una mayor cantidad de tiempo de escritura no necesariamente produce un impacto proporcional en la calidad de los escritos y, por el contrario, contribuye a ampliar la brecha en resultados entre estudiantes con mayor o menor formación y competencia escrita (APPLEBEE, 2007). En este sentido, un mayor impacto en la calidad de los escritos no estaría determinado por el tiempo sino más bien por las oportunidades que brinde el constructo a evaluar y la tarea, por ejemplo mediante la inclusión de espacios de planificación y preguntas orientadoras para la revisión (APPLEBEE, 2007).

En quinto lugar, y como consecuencia lógica de una definición actualizada de la escritura, las pruebas estandarizadas deberían tender al **uso de rúbricas holísticas para evaluar textos en contexto**. Existen dos diseños principales de medición de desempeños: rúbricas analíticas y rúbricas holísticas. Las rúbricas analíticas suelen incluir diversas dimensiones (por ejemplo: contenido, ortografía, léxico, cohesión, coherencia, estructura), cada una de ellas con distintos niveles de desempeño; frecuentemente, en pruebas de escritura se utilizan entre tres y cinco niveles. Una de sus fortalezas es que permiten otorgar información desagregada respecto de las dimensiones de escritura evaluadas. Sin embargo, las rúbricas analíticas utilizadas para medir textos complejos requieren mucho tiempo de evaluación (APPLEBEE, 2007) y resulta difícil lograr un nivel elevado de confiabilidad inter-analistas (PERELMAN, 2018). Por otro lado, es complejo determinar cuál es el peso específico de cada dimensión en el puntaje global obtenido desde la teoría de la escritura (¿cómo ponderar la estructura frente a la cohesión, o el léxico frente a la ortografía?) o de qué manera cuantificar cada una de esas dimensiones (¿cuántos errores ortográficos corresponden a cada nivel de desempeño?), pese a que técnicas como el análisis factorial o la normalización de puntajes de errores permitirían contestar estas preguntas desde la perspectiva de la medición. Resulta dudoso, por tanto, que las rúbricas analíticas arrojen información psicométricamente útil (APPLEBEE, 2007), afectando, como se expone en la próxima sección, la validez del instrumento. De hecho, existe escaso acuerdo respecto de cuáles son exactamente las dimensiones independientes que componen la competencia de escritura, y por eso su utilidad para

brindar diagnósticos y perfiles específicos de escritores ha sido vista con escepticismo (WHITE, 1984).

En términos más amplios, en la vida real los textos no funcionan de forma analítica sino como un producto articulado entre sus múltiples dimensiones. Tal como se afirma en una de las primeras propuestas de evaluación holística de la escritura en los años 60, desarrolladas en el *Educational Testing Service* (ETS), “to the extent that a unique communication has been created, the elements are related to the whole in a fashion that makes a high relationship of the parts inevitable. The evaluation of the part cannot be made apart from its relationship to the whole.” (ELLIOT, 2005, p. 160). En los años 80s, un artículo clave para la instalación de las rúbricas holísticas en los estudios de la escritura denunciaba el “reduccionismo analítico” como la imposición de una lógica de los artefactos a fenómenos humanos y sociales: “Analytic reductionism assumes that knowledge of the parts will lead to understanding of the whole, a theory which works very well with machinery or other objects, but less well with art forms or life forms.” (WHITE, 1984, p. 400). Por este motivo, la adecuación a contextos y el logro de propósitos no se puede ponderar evaluando dimensiones que no existen por separado. Como plantea White, la postura holística sobre la escritura rechaza la idea de que el todo es solo la suma de las partes. Además, algunos elementos de la escritura operan en múltiples dimensiones al mismo tiempo; por ejemplo, la elección léxica puede influir simultáneamente en otras dimensiones como el contenido técnico de lo afirmado, la construcción de lazos cohesivos con otros fragmentos del texto, el cumplimiento del propósito comunicativo o el establecimiento de vínculos con el lector. Por último, la evaluación holística ofrece un constructo que coloca el foco en las fortalezas del escrito, más que en sus deficiencias (CUSHING WEIGLE, 2002), y que resulta más cercano al funcionamiento de los textos y a la experiencia de los lectores, y por tanto tiene mayor validez: “Readers do not count the number of paragraphs or the number of sentences in a paragraph; they care if the text is complete, informative, and compelling.” (PERELMAN, 2018, p. 15).

En términos psicométricos, logísticos y económicos, las rúbricas holísticas garantizan una mayor confiabilidad entre jueces que las rúbricas analíticas (PERELMAN, 2018), permiten ganar en velocidad porque la evaluación surge de una sola lectura relativamente rápida (CUSHING WEIGLE, 2002) y ofrecen ventajas de costo-beneficio (WHITE, 1984).

En sexto lugar, los administradores de pruebas estandarizadas de escritura deben **socializar y consensuar el instrumento con la comunidad educativa**. La construcción contemporánea de instrumentos de medición de la escritura involucra la incorporación de paneles con miembros relevantes de la comunidad educativa, tales como profesores, directivos y académicos expertos en la temática (PERELMAN, 2018). Para O’Neill, Moore y Huot (2009) existe un desequilibrio de poder que ha llevado a que las llamadas “mediciones objetivas de la escritura” se perciban como medidas más exactas que el juicio de los profesores: “For at least one hundred years, then, reliable and valid writing assessment outside of a school context has been envisioned as a better source of evaluative information about students than teachers” (O’NEILL; MOORE; HUOT, 2009, p. 17). La incorporación de miembros diversos de la comunidad educativa permite determinar resultados de escritura **inclusivos** que eviten el establecimiento de una norma de escritura dominante y homogénea (POE; INOUE; ELLIOT, 2018). También es frecuente realizar investigaciones contextuales con los escritores que rinden el examen, indagando tanto en las relaciones entre resultados y variables sociodemográficas, como en sus percepciones y experiencias con el instrumento (estrategias, dificultades, conocimientos previos, etc.) recogidas mediante encuestas, entrevistas, grupos focales y protocolos de pensamiento en voz alta (CHENG; FOX; ZHENG, 2007). Por último, siguiendo principios contemporáneos de accesibilidad y transparencia (HUOT, 2009), los criterios de construcción de tareas y rúbricas, junto con muestras de escritura evaluadas, deben ofrecerse de forma pública a todos los interesados en las pruebas.

A partir de la sistematización del estado del arte de la evaluación en los estudios de la escritura, se propone un set de seis principios teóricos para la construcción de mediciones estandarizadas de escritura, como muestra la imagen 1, a continuación. Estas deben evaluar de manera directa muestras reales de escritura, solicitar tareas de escritura situadas y significativas, colocar a los estudiantes en posición de autoridad, autonomía y agencia, medir dimensiones escritas de alta complejidad, utilizar rúbricas holísticas apropiadas para textos que cumplen propósitos en contexto, y socializar y consensuar el instrumento con la comunidad educativa, considerando las perspectivas de los actores involucrados, particularmente docentes y estudiantes evaluados. Estos elementos, en conjunto, constituyen las bases teóricas para el desarrollo de pruebas de escritura válidas, entendiendo “validez” desde una mirada contemporánea de la evaluación, como se explica en profundidad en el siguiente apartado.

Imagen 1 - Principios teóricos para la validez de evaluaciones estandarizadas de escritura



Fuente: Los autores (2018).

4. Validez

Según O'Neil, Moore y Huot (2009), el concepto de "validez" ha estado en constante evolución desde su introducción en la década de 1930. Originalmente, validez se definió como un criterio de correlación de la prueba con otra medida objetiva, lo que posteriormente se conceptualizaría como "validez de criterio". Sin embargo, recién en la década de 1950, con la publicación de una primera versión de los estándares de medición educacional de American Educational Research Association (AERA), American Psychological Association (APA) y National Council on Measurement in Education (NCME), se realizó un esfuerzo sistemático por definir el concepto. La publicación de una nueva versión de este texto cada 15 años aproximadamente, la última de ellas en 2014 (AMERICAN EDUCATIONAL RESEARCH ASSOCIATION et al., 2014), da cuenta del carácter cambiante de este concepto.

En la versión de los años 60 de los estándares, por ejemplo, los tres focos centrales para la validación de pruebas correspondían a la **validez de contenido** (selección de aspectos para medir lo que se desea medir), de **criterio** (correlación de la prueba con otras medidas) y de **constructo** (conceptualización teórica de lo que se medirá). Sin embargo, la relativa juventud de los estudios de la escritura y la tendencia a sobredimensionar la importancia de la **confiabilidad** en la puntuación de las muestras escritas impidieron hasta hace muy poco la utilización de una definición teóricamente informada de escritura (O'NEIL; MOORE; HUOT, 2009). De este modo, por mucho tiempo el concepto de validez en la evaluación de la escritura estuvo confundido con el de confiabilidad, es decir, por el acuerdo alcanzado entre diferentes analistas para la medición del constructo, dada la complejidad de la puntuación de las muestras escritas (O'NEIL; MOORE; HUOT, 2009;

PERELMAN, 2018). Sin embargo, la confiabilidad no garantiza que el instrumento mida efectivamente el constructo de interés y no otros elementos, ni salvaguarda aspectos como los usos y consecuencias de la medición, su efecto y adecuación al contexto en que se efectúa o su cercanía con las características ecológicas del constructo medido (WILLIAMSON, 1993; MURPHY; YANCEY, 2008). Para lograrlo, se requiere que la definición de las variables medidas sea teóricamente sólida y fundamentada (CAMP, 2009), un aporte central al que aspira el presente trabajo.

Pero, además, una concepción contemporánea de validación de pruebas requiere que estas colaboren con la justicia (*fairness*) y uso apropiado del instrumento estandarizado de evaluación, en particular para apoyar las necesidades de los estudiantes (WIGGINS, 1990). Como afirman Luce-Kapler y Klinger, “high-stakes testing must be accompanied by explicit efforts to ensure the tests either support relevant educational goals or, at least, do not limit the educational domains being taught”. (LUCE-KAPLER; KLINGER, 2005, p. 171). De hecho, en publicaciones especializadas recientes, la consideración por los usos de los instrumentos de evaluación de la escritura (HUOT, 2009) ha derivado en discusiones sobre equidad, inclusión, ampliación de oportunidades de distintos grupos sociales y justicia social, en el marco de una perspectiva de **evaluación socialmente justa** de la escritura (*socially just writing assessment*; POE; INOUE; ELLIOT, 2018). En ocasiones, las pruebas de escritura operan con implícitos respecto de expectativas de desempeño que pueden resultar excluyentes para estudiantes menos familiarizados con las prácticas culturales y semióticas de los ámbitos académicos (CAMP, 2009). Por estos motivos, las interpretaciones, usos y consecuencias sociales, culturales y educativas de los instrumentos de evaluación han pasado a ocupar un rol central en la determinación de la validez (MURPHY; YANCEY, 2008). Como propone Broad, “we no longer seem to be content to be told that assessments meet certain psychometric and statistical requirement. We want to know whether they help or hurt learning and teaching.” (BROAD, 2003, p. 9).

La inclusión de tareas significativas y situadas de escritura, con escritores en roles de autoridad y agencia, en pruebas estandarizadas, de gran escala y de impacto en las trayectorias escolares destinadas a la educación superior podría fomentar la enseñanza intencionada de escritura en el sistema de educación formal. En sentido inverso, la ausencia de una prueba específica de escritura o la utilización de un constructo teóricamente mal informado representa un riesgo para los procesos de enseñanza y

aprendizaje de esa competencia (SLOMP, 2008). Este problema ya fue identificado, por ejemplo, en los años 60 por el *Educational Testing Service* (ETS) de Estados Unidos respecto de las pruebas de selección múltiple para medir de forma indirecta la competencia escrita (WHITE, 1984, p. 401), y ha generado resistencia y escepticismo de la comunidad de profesores a cargo de procesos de formación letrada (CAMP, 2009). Cabe aclarar que, si se omite la influencia negativa en la enseñanza y la poca validez de constructo de las pruebas indirectas de escritura mediante selección múltiple, existen evidencias de que el desempeño en este tipo de pruebas se corresponde bastante bien con el desempeño en pruebas de evaluación directa teóricamente informadas, aunque ambos instrumentos no están necesariamente evaluando las mismas habilidades (CAMP, 2009).

A su vez, la autenticidad, es decir, la producción de muestras directas de escritura individuales vinculadas a la consecución retórica de propósitos en contexto, propios de las tareas efectuadas en el contexto para el que se está haciendo la evaluación (MURPHY; YANCEY; 2008), puede favorecer políticas de diagnóstico y asignación de nivel más locales, en las instituciones en las que los estudiantes ingresen y a cargo de los propios profesores (HUOT, 2009), así como favorecer una mayor participación y uso democrático de los resultados en la comunidad. Como indica Wiggins, “authentic assessment also has the advantage of providing parents and community members with directly observable products and understandable evidences concerning their students’ performance” (WIGGINS, 1990, p. 3). Por último, la inclusión de temas y géneros discursivos diversos y optativos, con investigación previa de prácticas, necesidades y oportunidades semióticas y educativas, y con participación y en diálogo de miembros diversos de la comunidad, puede colaborar con la inclusión y valoración de trayectorias formativas y perfiles estudiantiles diversos.

En suma, la validez de una prueba estandarizada de escritura requiere considerar su fundamentación teórica y usos y consecuencias sociales y educativas. En la síntesis ofrecida por Camp, “all evidence for validity is to be interpreted in relation to the theoretical construct, the purpose for the assessment, and therefore the inferences derived from it, and the social consequences.” (CAMP, 2009, p. 116). La necesidad de proponer instrumentos justos y significativos, que den cuenta de los aspectos de la escritura que los profesores realmente valoran, y teóricamente fundamentados, informados por concepciones contemporáneas de la escritura, puede oponerse a la

tendencia a desarrollar instrumentos simples, rápidos, eficientes y de fácil acuerdo entre correctores (BROAD, 2003; CAMP, 2009; HUOT, 2009).

5. Evaluación de la escritura en perspectiva internacional: algunas prácticas observadas

Existen antecedentes a nivel regional e internacional de pruebas estandarizadas de gran escala que miden la competencia de escritura siguiendo algunos de los principios propuestos en este artículo. Sin ánimo de exhaustividad, se ofrecen a continuación algunos ejemplos de cómo estos se han implementado en la práctica internacional actual. En particular, todas las evaluaciones que se discuten en este apartado cumplen con el **Principio 1** de evaluar muestras reales de escritura, abandonando la medición indirecta.

Sobre el segundo principio, “solicitar tareas situadas y significativas”, estos instrumentos suelen definir de forma explícita una audiencia y un género discursivo específicos como parte de la tarea de escritura (PERELMAN, 2018). Por ejemplo, una tarea de escritura del *National Assessment of Educational Progress* (NAEP⁴) solicita a los estudiantes que escriban una carta al director del colegio sobre un tema específico, en un contexto particular, respecto del cual deben adoptar una posición⁵. En otro caso, una tarea del *Smarter Balanced Assessment Consortium*⁶ solicita a estudiantes de 6º grado la escritura de un artículo informativo sobre un tema específico para el periódico de la escuela que será leído por padres, profesores y estudiantes pares. Por su parte, los *English Language A Levels* (AQA⁷) solicitan géneros discursivos como el ensayo de opinión, mientras que la prueba de inglés avanzado como segunda lengua, *Proficiency*⁸, ofrece la posibilidad de elegir entre múltiples géneros como artículos, cartas, informes, reseñas y ensayos (CAMBRIDGE ENGLISH, 2016).

⁴ El National Assessment of Educational Progress (NAEP) es una evaluación periódica de competencias diversas, incluida la escritura, que desde 1969 se aplica en Estados Unidos a muestreos representativos de estudiantes de 4º, 8º y 12º grado.

⁵ “Some of your friends perform community service. For example, some tutor elementary school children and others clean up litter. They think helping the community is very important. But other friends of yours think community service takes too much time away from what they need or want to do. Your principal is deciding whether to require all students to perform community service. Write a letter to your principal in which you take a position on whether students should be required to perform community service. Support your position with examples.” (disponible en <https://nces.ed.gov/NationsReportCard/ngt/Search>; consultado el 22/10/2018).

⁶ El Smarter Balanced Assessment Consortium, creado en Estados Unidos en 2010, evalúa de forma articulada las competencias de lectura y escritura de estudiantes de 3º a 8º grado y de 11º grado en 13 estados de ese país.

⁷ Los English Language A Levels (AQA) son pruebas utilizadas en Reino Unido para medir diferentes competencias en la admisión a la educación superior.

⁸ El Cambridge English: Proficiency es una prueba estandarizada de gran escala que mide el dominio del inglés (británico) orientada a hablantes adultos de inglés como segunda lengua.

A nivel regional, la prueba TERCE⁹ considera la escritura como una manifestación cultural y, por ese motivo, “las consignas se diseñaron considerando el género carta (que se estudia mayoritariamente en los países participantes) y se adecuaron a situaciones comunicativas con un fin específico (acotado a secuencias textuales) y cercanas a niños y niñas de la región.” (FLOTTS et al., 2016, p. 12). En Chile, la prueba Simce¹⁰ Lenguaje y Comunicación: Escritura de 6° básico propone tareas de escritura en las que se especifica una situación, un género y un lector específico, ya que sigue un enfoque comunicativo en el que el lenguaje desempeña un rol en la construcción y comprensión del mundo y en el diálogo con uno mismo y con los demás (AGENCIA DE CALIDAD DE LA EDUCACIÓN, 2018a; 2018b).

Del mismo modo, estos instrumentos solicitan, en general, tareas de escritura distintas que deben ser resueltas en un tiempo mínimo apropiado para la actividad. Por ejemplo, el *National Assessment of Educational Progress* (NAEP) solicita la elaboración de dos textos durante 25-30 minutos cada uno. A su vez, se asignan tareas distintas de forma aleatoria, lo cual fomenta que haya igualdad de oportunidades para distintos grupos de estudiantes con distintos perfiles, trayectorias, experiencias y conocimientos. En pocos casos, como en el *Foundation Skills Assessment* (FSA)¹¹, los estudiantes pueden de hecho elegir las temáticas que prefieran. Por su parte, la prueba de escritura del *Scholastic Aptitude Test* (SAT)¹² solicita la escritura de un ensayo como respuesta a una lectura; el escritor cuenta con 50 minutos para resolver la tarea en su conjunto. En algunas pruebas, como la Simce escritura, los estudiantes cuentan con espacio para planificar, con sugerencias para hacerlo (“Lluvia de ideas. Haz un listado con todas las ideas que se te ocurran sobre lo que vas a escribir”), y preguntas que orientan la revisión del escrito (por ejemplo, “¿Seguiste la instrucción que se te dio?”). (AGENCIA DE CALIDAD DE LA EDUCACIÓN, 2018b, pp. 7-8). En el caso de pruebas de escritura en segundas lenguas, como el *Proficiency*, los escritores cuentan con 90 minutos para completar una tarea de

⁹ El Tercer Estudio Regional Comparativo y Explicativo (TERCE), desarrollado por el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), a su vez coordinado por la Oficina Regional de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (OREAL/UNESCO, Santiago), se aplicó en 2013 a estudiantes de 3° y 6° básico de 15 países de América Latina y el Caribe en las áreas de Lectura, Escritura, Ciencias y Matemáticas.

¹⁰ La prueba Simce Lenguaje y Comunicación: Escritura evalúa las habilidades de escritura de estudiantes de 6° básico de escuelas chilenas a partir de los Objetivos de Aprendizaje de las Bases Curriculares especificadas por el Ministerio de Educación (AGENCIA DE CALIDAD DE LA EDUCACIÓN, 2018a).

¹¹ El *Foundation Skills Assessment* (FSA) es una prueba que evalúa matemáticas, lectura y escritura en estudiantes de 4° y 7° grado de la provincia canadiense de British Columbia.

¹² El *Scholastic Aptitude Test* (SAT) es una prueba orientada a medir distintas competencias al final de la educación media; desde 2016, el SAT incluye la escritura opcional de un ensayo.

escritura preasignada no optativa y una segunda tarea optativa de entre cinco posibilidades (CAMBRIDGE ENGLISH, 2016).

Respecto de las temáticas propuestas, se busca que no existan sesgos por género, grupo social de pertenencia u otros factores, y que no resulten socialmente controvertidas. Estos criterios colaboran con la inclusión social en pruebas de amplio alcance en la población. En el caso del National Assessment of Educational Progress (NAEP), se coloca el foco en temas generales y de fácil acceso, a través de fragmentos cortos para leer y estímulos visuales, sobre experiencias vitales frecuentes, contenido académico general (libros y música favoritos, personajes históricos y ficticiales, etc.) y participación ciudadana (temas pertinentes para la comunidad o para el entorno escolar) (APPLEBEE, 2007).

Otro de los principios representados en la práctica actual es el **Principio 4** sobre la medición de dimensiones de alta complejidad. Varios de estos instrumentos ponderan especialmente aspectos vinculados a la adecuación a situaciones, lectores y géneros discursivos, frente a dimensiones normativas. La prueba Simce de escritura, por ejemplo, considera que “es más determinante, al momento de comunicar un mensaje con claridad, escribir de manera coherente que utilizar correctamente las normas de ortografía acentual.” (AGENCIA DE CALIDAD DE LA EDUCACIÓN, 2018a, p. 5). En el caso del *Proficiency*, se evalúan cuatro dimensiones: relevancia de contenidos; cumplimiento de propósitos comunicativos; cohesión y coherencia del texto; y dimensiones lingüísticas como vocabulario y gramática (CAMBRIDGE ENGLISH, 2016). Para las pruebas con foco en la continuación de estudios luego de la educación media, como el *National Assessment of Educational Progress* (NAEP), el énfasis se coloca en la evaluación de habilidades de escritura que tendrán continuidad en la educación superior, pero no necesariamente en las clases tipificadas de escritura propias de ese nivel educativo; algunas de esas habilidades, orientadas por los propósitos comunicativos y los destinatarios del texto, son desarrollo apropiado de ideas, organización lógica, fluidez lingüística, uso de convenciones, análisis efectivo, interpretación, resolución de problemas (APPLEBEE, 2007).

Con respecto al **Principio 5**, estos instrumentos efectivamente utilizan rúbricas holísticas (como el *National Assessment of Educational Progress* y el *ACT Scaling Test*¹³) u holísticas combinadas con evaluación analítica (como el *Foundation Skills Assessment*). En el caso de la prueba Simce de escritura, se utiliza una pauta holística que “permite evaluar los textos de los estudiantes desde una perspectiva global, ya que la descripción de cada nivel de esta pauta apunta a cómo la puesta en práctica de las habilidades en su conjunto permite transmitir mensajes con claridad según las diferentes situaciones comunicativas que se le presentan al estudiante.” (AGENCIA DE CALIDAD DE LA EDUCACIÓN, 2018a, p. 4); de forma complementaria se usa una pauta analítica para entregar información más detallada a los establecimientos educativos, velando por el uso educativo que puede darse a esta prueba. Respecto de exámenes de escritura en segundas lenguas, el TOEFL¹⁴ utiliza una rúbrica holística con 6 niveles de desempeño (CUSHING WEIGLE, 2002).

Como se puede apreciar en este breve recorrido, las pruebas, en general, cumplen varios de los principios teóricos para la validez formulados en este texto, aunque con algunas debilidades, en particular relacionadas con la falta de autenticidad de las tareas al usar géneros que no tienen circulación real en las aulas o géneros discursivos de baja especificidad como el “ensayo”, lo que atenta contra el **Principio 2**, solicitar tareas situadas y significativas. En particular, algunas pruebas fallan en un aspecto teórico que atenta directamente contra su validez: no definen con claridad un propósito comunicativo y retórico. En otras palabras, el propósito de los textos solicitados corresponde a una tipología textual encubierta como el propósito de “narrar”, el propósito de “explicar” o el propósito de “informar”. Se trata de abstracciones que no reflejan el carácter de acción en el mundo propio de una definición actualizada de la escritura como la que se ofrece al inicio de este artículo.

Por otro lado, el **Principio 3** sobre ubicar a los estudiantes en posición de autoridad, autonomía y agencia se ve ausente al constatar la imposibilidad de que los estudiantes escojan los temas o el uso de temas con sesgos culturales o de género. Con todo, es

¹³ El ACT Scaling Test es una prueba administrada en el Distrito Federal de Australia a estudiantes de 12º grado, es decir, en su último año de educación media obligatoria.

¹⁴ El Test of English as a Foreign Language (TOEFL) es una prueba estandarizada de gran escala que mide el dominio del inglés (estadounidense) orientada a hablantes adultos de inglés como segunda lengua. Según la rúbrica holística que evalúa la producción escrita, un ensayo en el nivel máximo de desempeño “effectively addresses the writing task, is well organized and well developed, uses clearly appropriate details to support a thesis or illustrate ideas, displays consistent facility in use of language, demonstrates syntactic variety and appropriate word choice though it may have occasional errors” (CUSHING WEIGLE, 2002, p. 113).

importante resaltar aquellos aspectos que sí ejemplifican la factibilidad de avanzar hacia una evaluación de la escritura mejor informada teóricamente.

6. Evaluación de la escritura en la Prueba de Selección Universitaria (PSU)

En contraste, algunas pruebas de alto impacto social, cultural y educativo en la región utilizan un constructo no informado en múltiples niveles y muchas veces no siguen criterios claros de justicia social y uso pedagógicamente provechoso de la evaluación, como en el caso del componente de escritura de la Prueba de Selección Universitaria (PSU) de Chile. La PSU tiene como propósito general la selección de postulantes a la educación superior chilena a partir de la medición de habilidades cognitivas y de contenidos. Se basa en los objetivos fundamentales y los contenidos mínimos obligatorios del currículum en cuatro áreas: Lenguaje y Comunicación; Matemática; Historia, Geografía y Ciencias Sociales; Ciencias Naturales (DEPARTAMENTO DE EVALUACIÓN, MEDICIÓN Y REGISTRO EDUCACIONAL, 2018a). A lo largo de sus 15 años de existencia, ha sido objeto de polémicas. En 2004, en su primer año de implementación, el *Educational Testing Service* (ETS) recomendó que se investigaran los usos de la prueba más allá de su rol en los procesos de selectividad (PEARSON, 2013). En 2013, Pearson realizó sugerencias sobre el sesgo de la prueba respecto del currículum escolar de los establecimientos de origen (humanista-científico o técnico-profesional) y de ciertos grupos sociales poco representados; además, enfatizó la importancia de que "the theoretical frameworks and the specifications of the tests be submitted for validation by readers from outside the committee members." (PEARSON, 2013, p. 12), así como la recolección mediante encuestas de perspectivas de estudiantes evaluados.

Como parte del componente de Lenguaje, la Prueba de Selección Universitaria incluye la evaluación indirecta de la escritura. Fundamentada en el currículum escolar, busca "valorar la escritura como una actividad creativa y reflexiva de expresión personal, que permite organizar las ideas, presentar información, interactuar con diversas realidades y como una oportunidad para elaborar conscientemente una visión personal del mundo", fomentar el "escribir, utilizando flexible y creativamente, la variedad de recursos que ofrece la lengua, de acuerdo con la estructura y el sentido semántico del texto" y dar cuenta de la "aplicación del proceso general de escritura (planificación, escritura, revisión, reescritura, edición)", entre otros elementos de enseñanza obligatoria (DEPARTAMENTO DE EVALUACIÓN, MEDICIÓN Y REGISTRO EDUCACIONAL, 2018a, p. 4).

Sin embargo, el instrumento de evaluación no responde a estos propósitos ni mide estas dimensiones. En su primera parte, "Manejo de conectores", los estudiantes deben seleccionar una opción de entre cinco (selección múltiple cerrada) que presenta uno o dos conectores para diez oraciones aisladas. Se aprecia aquí una teoría implícita de la escritura como disponibilidad léxica descontextualizada. En su segunda parte, "Plan de redacción", los estudiantes deben ordenar una secuencia de cinco enunciados para componer un texto coherente, como se aprecia en la imagen 2 a continuación:

Imagen 2 – Actividad del Plan de Redacción.

- "Los pimientos"
1. Entre los más conocidos se encuentran el pimiento morrón, la guindilla y el chile.
 2. Los pimientos son ricos en vitamina C, vitamina A y licopeno, sobre todo el pimiento rojo.
 3. Pertenece a la familia de las solanáceas, al igual que la papa y el tomate.
 4. La pimentera (*capsicum annum*) es una planta procedente de América, cuyo cultivo se extiende por todo el mundo.
 5. El pimentón y los pimientos picantes poseen propiedades analgésicas, por lo que son especialmente recomendados para tratar el reumatismo y las neuralgias.

Fuente: DEPARTAMENTO DE EVALUACIÓN, MEDICIÓN Y REGISTRO EDUCACIONAL (2018c, p. 7).

Esta tarea de escritura supone que los escritos tienen una estructura fija (que en este caso es 4-3-2-5-1), si bien podrían proponerse otras estructuras alternativas (aquí 4-3-2-1-5, no ofrecida como opción en la prueba), lo cual revela arbitrariedad en el diseño. La tarea se centra en medir la capacidad del estudiante para hallar indicadores semánticos y morfológicos que le permitan ordenar la secuencia de un texto preescrito, una competencia que no forma parte de los usos reales de la producción escrita. En términos más amplios, ninguno de los componentes de la prueba sigue una noción de la escritura como actividad compleja, significativa, diversa, tipificada en géneros discursivos, y situada. No se ofrecen posibilidades de autoría, agencia o creatividad para los estudiantes, ni se evalúan de manera directa muestras reales de escritura.

Este tipo de críticas ha derivado en la necesidad de generar un instrumento estandarizado que efectivamente mida la escritura de forma directa, válida y justa en el ingreso a la educación superior chilena. Esta propuesta fue posible en el marco del desarrollo más amplio de una batería de instrumentos para la evaluación de competencias fundamentales y habilidades transversales necesarias para cursar exitosamente un programa de estudios en el sistema de educación superior, dentro del

proyecto financiado ID16I10090 del Fondo de Fomento al Desarrollo Científico y Tecnológico (FONDEF) de Chile.

7. Metodología de la elaboración de la nueva propuesta evaluativa

En este apartado se reporta la cronología de acciones y decisiones que llevó al desarrollo del instrumento que se presenta a continuación.

- La primera fase del proyecto consistió en la identificación, a partir de la revisión de la literatura, de los seis **Principios teóricos para la validez de evaluaciones estandarizadas de escritura** presentados en este texto, en el marco de grupos de estudio y sesiones de discusión entre los autores y los especialistas encargados de construir y administrar la actual Prueba de Selección Universitaria (PSU) de Lenguaje.
- La siguiente acción consistió en la elaboración de un constructo de escritura a ser evaluado, teóricamente informado y que diera cuenta de las demandas de la escritura propias de la etapa en que sería administrada la prueba: la transición desde la escuela a la educación superior (ver en siguiente apartado), asumiendo que el nuevo instrumento, de acuerdo con el Principio 1, evaluará muestras directas de escritura.
- Durante el primer semestre del año 2016 se elaboró un modelo de dos tareas de escritura (“entrada de enciclopedia” y “transformación de infografía”) que permitieran dar cuenta del constructo elaborado y que cumplieran con las exigencias de los Principios 2 y 3. Estos corresponden al carácter situado y significativo, por medio de la escritura auténtica de un género conocido, y a la posición de autoridad y agencia, mediante la elección de temáticas y actividades en las que los estudiantes escritores sean competentes.
- Para estas tareas diseñadas, el Departamento de Evaluación, Medición y Registro Educativo de la Universidad de Chile (DEMRE) implementó una aplicación piloto entre el 22 de agosto y el 5 de septiembre de 2016 en 22 establecimientos educacionales humanístico-científicos y técnico-profesionales, de tres dependencias (municipales, particulares pagados y particulares subvencionados), pertenecientes a las comunas de Santiago, Talagante, Colina y Las Condes. Para el piloto se convocaron 2230 estudiantes de último año de secundaria (grado 12º) que completaron dos tareas. Esta experiencia permitió constatar el buen

desempeño de una de las tareas (“entrada de enciclopedia”) e identificar debilidades en la otra (“transformación de infografía”). Con las muestras de escritura obtenidas, se construyeron *benchmarks* en 4 niveles de desempeño y se procedió a la elaboración de una rúbrica que cumpliera con los Principios 4 y 5: medir dimensiones de alta complejidad y utilizar rúbricas holísticas centradas en el propósito.

- Con la información disponible y una primera propuesta de rúbricas de evaluación, un equipo de constructores llevó a cabo un proceso de elaboración de nuevas tareas de escritura durante el primer semestre de 2017 a cargo del DEMRE.
- Por último, entre el 25 de julio y el 3 de agosto de 2017 se efectuó una aplicación piloto a 2141 estudiantes de último año de secundaria (grado 12º) procedentes de 56 establecimientos educacionales de tres dependencias (municipales, particulares pagados y particulares subvencionados), pertenecientes a tres zonas geográficas: Zona norte caracterizada por la IV región de Coquimbo; Zona centro caracterizada por la región Metropolitana y la VI región de O’Higgins; Zona sur caracterizada por la VIII región del Bío-Bío. Se trata de una muestra no representativa de la población (N=117.298), pero con un margen de error >3%, lo que permite suponer una alta probabilidad de hacer inferencias pertinentes, aun cuando no se buscó potencia estadística en la constitución de la muestra. Se aplicaron dos formas de prueba con dos tareas cada una, incluida la “entrada de enciclopedia”.
- Posteriormente, en marzo de 2018, se realizó una aplicación piloto de una prueba con dos tareas, incluida la “entrada de enciclopedia”, a 857 estudiantes ingresantes (de un total de 932 matriculados) a las 16 carreras de una universidad chilena pública, regional y de selectividad media. Los resultados del comportamiento de la tarea “entrada de enciclopedia” en los pilotos de 2017 y 2018 se reportan en el siguiente apartado.

A la fecha, el instrumento continúa en fases de ajustes por parte del DEMRE, con el fin de cumplir con los parámetros psicométricos necesarios para escalarlo a una aplicación nacional. Entre el 6 y 17 de agosto de 2018, se aplicó un último piloto con dos formas de prueba con dos tareas cada una a 832 estudiantes del último año de secundaria, provenientes de 23 establecimientos educacionales de tres dependencias (municipales, particulares pagados y particulares subvencionados), pertenecientes a las regiones

Metropolitana y de O'Higgins de Chile. Actualmente se está implementando un proceso de corrección tendiente a estabilizar los niveles de la rúbrica y los procedimientos de revisión.

8. Resultados: hacia una nueva evaluación de la escritura al inicio de la educación superior

A continuación, se ofrece una propuesta original de medición estandarizada de la competencia de escritura en el ingreso a la educación superior que se ha desarrollado desde el año 2016 por el Departamento de Evaluación, Medición y Registro Educacional, siguiendo los criterios previos de validez.

8.1 Definición de competencia

El constructo se basa en la noción general de competencia. Se diferencia del concepto de **conocimiento**, definido como la representación mental de hechos, principios y teorías de un dominio particular (WESTERA, 2001), según el cual la escritura solo constituiría un medio de acceso a esta representación (VERA, 2015); y también se diferencia de la noción de **habilidad**, definida como operación cognitiva de procesamiento (capacidad de síntesis, de interpretación textual, etc.) que permite realizar actos (DAVIES et al., 1999), según la cual la escritura permitiría hacer inferencias sobre la habilidad lingüística de los sujetos (BACHMAN, 1990). La noción de competencia, en cambio, puede entenderse como una evolución del modelo en base a habilidades que subraya la necesidad de establecer un vínculo directo entre la habilidad y su aplicación en situaciones reales. Según Pellegrino y Hilton (2012), la noción de competencia permite dar cuenta del carácter articulado e interdependiente de habilidades y conocimientos en ámbitos de saber y de acción específicos. En el nivel superior, el Consejo de Rectores de las Universidades Chilenas propone adoptar el modelo de diseño por competencias en los procesos de innovación curricular universitaria (CRUCH, 2013). Mientras que el modelo de habilidades supone que estas son generalizables, el constructo competencia está anclado a tareas específicas que se consideran relevantes y pertinentes de acuerdo con los objetivos de la evaluación (KLIEME; HARTIG; RAUCH, 2008).

Para la construcción teóricamente validada de la **competencia de escritura** se incluye la función epistémica (construcción de conocimiento) y semiótica (transmisión de conocimiento) de la escritura. A su vez, estas dos funciones generales de la escritura se

especifican en propósitos comunicativos concretos, que el texto busca cumplir según el género discursivo del que se trate, y en la adecuación a una situación comunicativa y retórica específica y a lectores particulares, con los que se establecen relaciones diversas (de simetría o asimetría de poder, de proximidad o lejanía, etc.), en marcos escolares, culturales e históricos más amplios. Además, el texto resultante debe guardar coherencia en el desarrollo y estructuración de las ideas y establecer lazos cohesivos entre sus partes, incluyendo recursos tanto léxicos como gramaticales. Por último, se debe ajustar el texto a las convenciones lingüísticas del español escrito y formal propio de entornos académicos y ciudadanos relevantes para el género discursivo solicitado, tanto en sus dimensiones gramaticales como de puntuación y ortográficas. A continuación, se ofrece la definición de la competencia de escritura:

Competencia de escritura: capacidad de construir y comunicar conocimiento por escrito para cumplir con propósitos específicos y adecuarse a una situación comunicativa particular a través de un texto coherente, cohesivo y ajustado a las convenciones lingüísticas del español. (DEPARTAMENTO DE EVALUACIÓN, MEDICIÓN Y REGISTRO EDUCACIONAL, 2018b).

Los propósitos, destinatarios, temáticas y situaciones comunicativas de las tareas de escritura propuestas se especifican en géneros discursivos vinculados a ámbitos de práctica y participación tanto escolares y académicos como ciudadanos. De esta manera, para las tareas de escritura se han desarrollado géneros discursivos como una reseña sobre un producto cultural (libro o serie) destinada al diario escolar; una carta de solicitud a una autoridad política local; o una entrada de enciclopedia sobre tecnología destinada a adultos mayores (ver más adelante).

8.2 Estructura de la tarea de escritura

En la prueba, se solicitan dos tareas de escritura asociadas a dos géneros discursivos distintos. Si bien es posible que el estudiante no haya escrito antes un ejemplar del género, o incluso que no lo haya leído con anterioridad, la tarea lo orienta de forma detallada respecto de las expectativas de escritura. Primero, incluye una descripción de la situación comunicativa o retórica en la que se enmarca la tarea solicitada. A continuación, se indica la tarea a llevar a cabo, con instrucciones precisas respecto de los propósitos comunicativos, del contexto y de la audiencia. Posteriormente, se entrega información respecto de las expectativas de evaluación del género discursivo solicitado.

En la imagen 3 puede apreciarse la tarea de escritura correspondiente a la elaboración de una entrada de enciclopedia:

Imagen 3 - Tarea de escritura de entrada de enciclopedia.

TEXTO 1

Instrucciones

Estás colaborando con un programa en el que se enseña a los adultos mayores a usar nuevas tecnologías (alfabetización digital). Te han solicitado crear **entradas de enciclopedia**, como las de Wikipedia, para explicar a los participantes qué son y cómo funcionan diferentes equipos tecnológicos, aplicaciones o redes sociales, por ejemplo, un teléfono inteligente o smartphone, una aplicación como Gmail o una red social como Facebook, entre otros.

Escoge un equipo, aplicación o red social y escribe una entrada de enciclopedia, como las de Wikipedia. Para ello, debes desarrollar más de un aspecto sobre el equipo, aplicación o red social, desde una perspectiva neutral y objetiva. Recuerda que tu texto debe ser comprensible para un adulto mayor.



Importante

Una entrada de enciclopedia es un texto que define un tema de manera más extensa que en un diccionario y con una perspectiva objetiva. En la entrada de enciclopedia se incluye información sobre el tema, por ejemplo: qué es, qué características tiene, cómo se usa, para qué sirve, cómo ha cambiado en el tiempo, etcétera.

Fuente: DEPARTAMENTO DE EVALUACIÓN, MEDICIÓN Y REGISTRO EDUCACIONAL (2018b).

A su vez, si bien la lectura y la escritura (e incluso la oralidad) suelen articularse en el uso efectivo del lenguaje en contexto, para poder evaluar específicamente el nivel de logro de la competencia de escritura no se incluyen lecturas previas. Por el contrario, se ofrece un conjunto de temáticas abiertas entre las cuales el estudiante puede elegir. De esta manera, el instrumento resulta más justo e inclusivo respecto de los conocimientos, experiencias y trayectorias diversas de los estudiantes, y ofrece oportunidades de agencia, autonomía y autoridad en la posición de enunciación del escritor. La solicitud de textos que cumplen propósitos académicos y ciudadanos relevantes y que apelan a saberes de los estudiantes colabora con la pertinencia y significancia de la tarea.

Durante la evaluación, el estudiante cuenta con 80 minutos para resolver las dos tareas de escritura, en el orden y con la distribución de tiempo que prefiera. Se espera que elabore dos escritos de alrededor de 350 palabras cada uno. Para orientar y modelar el proceso de escritura, además de una tarea explícita respecto de la situación, propósitos y expectativas del género, se ofrece un espacio para la planificación del texto (de forma escrita o multimodal), que no es evaluado. Además, se ofrece un conjunto de indicaciones finales breves y específicas (preguntas de cotejo) para intencionar la revisión del producto escrito, con referencia a los aspectos centrales de evaluación que ya fueron anticipados en la tarea.

8.3 Rúbrica

Para la corrección de los productos escritos, se utilizan rúbricas holísticas con cuatro niveles de desempeño que dan cuenta del funcionamiento integrado de la escritura como competencia compleja; que permiten evaluar en general su grado de adecuación a situaciones y cumplimiento de propósitos; y que favorecen la consistencia entre correctores (ver más arriba). En estos niveles de desempeño se articulan diferentes dimensiones textuales interdependientes, vinculadas con las dimensiones incluidas en la competencia de escritura: 1) grado de cumplimiento de propósitos específicos y de adecuación a una situación comunicativa; 2) coherencia y cohesión; 3) ajuste a convenciones lingüísticas del español. Estas tres dimensiones se especifican según los rasgos característicos de cada género discursivo.

Para construir la definición de los niveles de desempeño, se clasifican los resultados de los sucesivos piloteos de las pruebas en *benchmarks*, se identifican categorías discursivas y lingüísticas emergentes que permitan distinguir entre niveles, y se elaboran rúbricas iniciales. La imagen 4 representa un ejemplo de un *benchmark* del nivel más alto (4) y la imagen 5 un ejemplo de un *benchmark* del nivel más bajo de desempeño (1), en la primera versión de la rúbrica.

Imagen 4 - Ejemplo de desempeño nivel máximo.

Facebook

Es una aplicación y red social de personas.

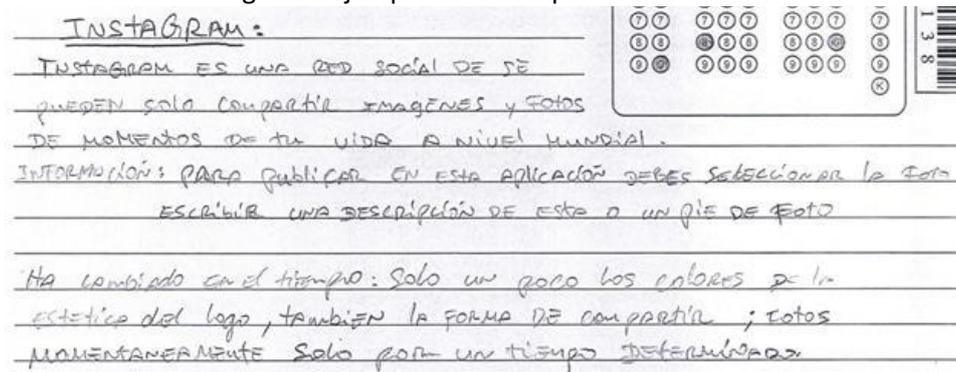
Sus características son que se la funciona con internet, es una red social destinada a cualquier tipo de personas ya sea jóvenes, adultos o personas mayores, tiene distintos tipos de uso y fines. Cuenta con juegos, chat, búsqueda de personas, páginas, personajes, etc, es una red social de carácter mundial.

Su uso es a través ya sea de un computador, tablet, teléfono, Smartphone, etc. Para poder utilizar esta aplicación es necesario crear una cuenta a través de un formulario de datos que se debe llenar. Su uso es muy sencilla una vez que la cuenta esta creada, a través de un tutorial el usuario se le va enseñando paso a paso como es el uso y manejo y las distintas acciones y usos con los que cuenta esta aplicación.

Esta red social cuenta con muchas usos y sirve para varias cosas entre las más importantes de puede buscar personas que hace tiempo no se veían, crear páginas gratuitas y jugar distintos tipos de juegos. Esta red a tenido un cambio importantísimo a través del tiempo tanto en su forma estructural como en su variedad de usos.

Fuente: DEPARTAMENTO DE EVALUACIÓN, MEDICIÓN Y REGISTRO EDUCACIONAL (2018b).

Imagen 5 - Ejemplo de desempeño nivel mínimo



Fuente: DEPARTAMENTO DE EVALUACIÓN, MEDICIÓN Y REGISTRO EDUCACIONAL (2018b).

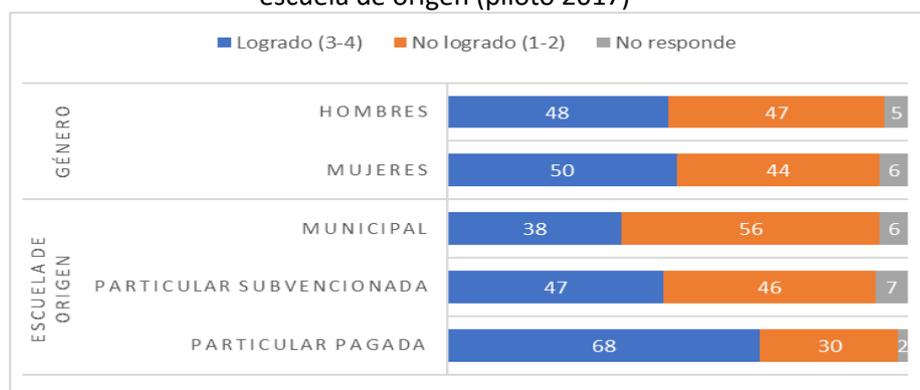
Posteriormente, estas rúbricas son piloteadas y validadas por expertos externos, quienes aportan sugerencias de ajustes y cambios. Para la corrección, se entrena a un conjunto de correctores a partir de muestras prototípicas de cada nivel de desempeño. Los ejemplares son corregidos de forma doble ciega; un sistema informático informa cuando aparecen inconsistencias entre correctores y se discuten los criterios utilizados. En la actualidad, se está trabajando en establecer un proceso de corrección de mayor sofisticación ya que hasta ahora no se ha alcanzado un índice de acuerdo lo suficientemente robusto. Se está testeando ampliar la rúbrica a seis niveles de desempeño, mejorar los procedimientos y el proceso de selección y entrenamiento de los correctores, y comenzar los pasos de corrección con una distinción entre textos logrados (desempeños superiores) y no logrados (desempeños inferiores) antes de la determinación del nivel específico al que corresponde cada texto.

8.4 Comportamiento de la tarea

En 2017 se realizó un estudio piloto de la tarea **entrada de enciclopedia** con 2141 estudiantes del último año de la educación media obligatoria chilena. Los resultados mostraron un porcentaje elevado de participación e interés de los estudiantes, ya que el 90% respondió la tarea. Este porcentaje es especialmente llamativo dado que se trató de una propuesta de evaluación y de producción escrita con la que los estudiantes no estaban familiarizados y que no estaba asociada a una calificación o a una obligación escolar. A su vez, los niveles de desempeño arrojaron una distribución normal, aunque con tendencia a concentrarse en los niveles medios: no logrado = 46% (nivel 1 **muy bajo** = 5%; nivel 2 **bajo** = 41%); logrado = 49% (nivel 3 **alto** = 40%; nivel 4 **muy alto** = 9%); respuestas que no respondieron la tarea = 5%. En términos sociodemográficos, se

identificaron diferencias según la escuela de origen y las similitudes por género estudiantil, resultados que indican un comportamiento esperado del instrumento, como resume la imagen 5:

Imagen 5 – Nivel de desempeño en escritura de entrada de enciclopedia en % por género y escuela de origen (piloto 2017)



Fuente: Los autores (2018) adaptado de DEPARTAMENTO DE EVALUACIÓN, MEDICIÓN Y REGISTRO EDUCACIONAL (2018b).

En marzo de 2018 se realizó un segundo estudio piloto con 857 estudiantes ingresantes (de un total de 932 matriculados) a las 16 carreras de una universidad chilena pública, regional y de selectividad media. En este caso, la distribución tuvo mayor concentración en el nivel 2 de desempeño, y una casi nula presencia del nivel más bajo: no logrado = 64% (nivel 1 **muy bajo** = 1%; nivel 2 **bajo** = 63%); logrado = 23% (nivel 3 **alto** = 18%; nivel 4 **muy alto** = 5%); respuestas que no respondieron la tarea = 13%. A su vez, los niveles de desempeño mostraron variaciones por carrera: aquellas con más del 30% de ingresantes con un nivel de desempeño **logrado** (nivel 3 y 4) fueron Medicina, Psicología, Derecho y Pedagogía en Inglés. Por el contrario, las carreras con más del 70% de ingresantes con un nivel de desempeño **no logrado** (nivel 1 y 2) fueron Administración, Pedagogía en Educación Especial, Veterinaria, Pedagogía en Lenguaje, Pedagogía en Educación Básica e Ingeniería Agronómica.

9. Conclusiones

Si se sigue una concepción actualizada de la escritura como tecnología epistémica y semiótica compleja, significativa, diversa, tipificada, situada y transversal, si se busca que la medición de los desempeños de los estudiantes en escritura redunde en más posibilidades de inclusión y equidad y en mejores prácticas de enseñanza y procesos de aprendizaje, entonces se necesita seguir una mirada contemporánea de validez para

pruebas estandarizadas de escritura. Ya no es aceptable la falta de diálogo y colaboración entre los expertos del campo de los estudios de la escritura, por un lado, y las instituciones encargadas de desarrollar instrumentos de medición, por el otro. Una medición estandarizada de la escritura debe evaluar de manera directa muestras reales de escritura, solicitar tareas de escritura situadas y significativas, colocar a los estudiantes en posición de autoridad, autonomía y agencia, medir dimensiones escritas de alta complejidad, utilizar rúbricas holísticas apropiadas para textos que cumplen propósitos en contexto, y socializar y consensuar el instrumento con la comunidad educativa, tal como sucede con el test para la evaluación de la escritura en el ingreso a la educación superior presentado aquí. De esta manera, la validez debe desplazarse desde aspectos psicométricos clásicos como la confiabilidad inter-correctores hacia la fundamentación teórica y la justicia de los usos y consecuencias de las mediciones, de forma tal de ofrecer pruebas de escritura verdaderamente significativas para la sociedad.

En este artículo se ha ofrecido un marco sistemático que se hace cargo de estos principios teóricos para la elaboración de pruebas de escritura válidas tanto desde el punto de vista de su fundamentación teórica como desde su uso e impacto social, sin desconocer, además, el criterio clásico de confiabilidad y los aspectos de tipo logístico y metodológico. Este marco, utilizado de manera creciente en pruebas estandarizadas a gran escala internacionales y regionales de los últimos años, ha permitido la elaboración y el piloteo de una propuesta original de medición estandarizada de escritura en el ingreso a la educación superior chilena, con resultados preliminares normales de desempeño y evidencias de participación e interés. Este instrumento y el marco teórico que lo sustenta permiten avizorar un cambio posible en los instrumentos actuales de medición en Chile y en otros países de la región.

REFERENCIAS

- ADLER-KASSNER, L.; WARDLE, E. (Ed.). *Naming what we know: threshold concepts of writing studies*. Logan, Utah: Utah State University Press, 2015.
- AGENCIA DE CALIDAD DE LA EDUCACIÓN. *Marco de evaluación prueba simce lenguaje y comunicación: escritura 6° año básico*. Santiago, 2018a. Disponible en: <http://archivos.agenciaeducacion.cl/Marco_de_evaluacion_simce_lenguaje.pdf>. Consultado el: 22 oct. 2018.
- _____. *Ejemplos de ítems texto informativo escritura: 6° básico* 2016. Santiago, 2018b. Disponible en: <http://archivos.agenciaeducacion.cl/2_Ejemplos_de_items_texto_informativo.pdf>. Consultado el: 22 oct. 2018.
- AMERICAN EDUCATIONAL RESEARCH ASSOCIATION et al. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, 2014.
- APPLEBEE, A. N. Issues in large-scale writing assessment: perspectives from the national assessment of educational progress. *Journal of Writing Assessment*, [S.l.], v. 3, n. 2, p. 81-98, 2007.
- BACHMAN, L. F. *Fundamental considerations in language testing*. Oxford: Oxford University Press, 1990.
- BAJTÍN, M. M. El problema de los géneros discursivos. In: _____. *Estética de la creación verbal*. Buenos Aires: Siglo XXI Editores, 2005. p. 248-293.
- BARTON, D.; HAMILTON, M. Literacy practices. In: BARTON, D.; HAMILTON, M.; IVANIČ, R. (Ed.). *Situated literacies- reading and writing in context*. New York: Routledge, 2000. p. 7-15.
- BAZERMAN, C. Understanding the lifelong journey of writing development. *Infancia y Aprendizaje: Journal for the Study of Education and Development*, [S.l.], v. 36, n. 4, 421-441, 2013.
- BROAD, B. *What we really value: beyond rubrics in teaching and assessing writing*. Logan, Utah: Utah State University Press, 2003.
- CAMBRIDGE ENGLISH. Cambridge English proficiency: handbook for teachers for exams from 2016. Cambridge: Cambridge University Press, 2016. Disponible en: <<https://www.cambridgeenglish.org/images/168194-cambridge-english-proficiency-teachers-handbook.pdf>>. Consultado el: 22 oct. 2018.
- CAMP, R. Changing the model for the direct assessment of writing. In: HUOT, B.; O'NEILL, P. (Ed.). *Assessing writing: a critical sourcebook*. Boston: Bedford & St. Martin's, 2009. p. 102-129.
- CHENG, L.; FOX, J.; ZHENG, Y. Student accounts of the Ontario secondary school literacy test: a case for validation. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, Canadá, v. 64, n. 1, p. 69-98, 2007.

- CHIROLEU, A.; MARQUINA, M. Democratisation or credentialism?: public policies of expansion of higher education in Latin America. *Policy Reviews in Higher Education*, [S.l.], v. 1, n. 2, p. 139-160, 2017.
- CRUCH. *Manual para la implementación del sistema de créditos académicos transferibles SCT-Chile*. Santiago, 2013.
- CUSHING WEIGLE, S. *Assessing writing*. Cambridge: Cambridge University Press, 2002.
- DAVIES, A. et al. *Dictionary of language testing*. Cambridge: Cambridge University Press, 1999.
- DEPARTAMENTO DE EVALUACIÓN, MEDICIÓN Y REGISTRO EDUCACIONAL. *Temario PSU: lenguaje y comunicación*. Santiago: DEMRE, 2018a. Disponible en: <<http://www.psu.demre.cl/publicaciones/pdf/2019-18-04-12-temario-lenguaje.pdf>>. Consultado el: 22 oct. 2018.
- _____. *Prueba de escritura*. Santiago: DEMRE, 2018b. (No prelo).
- _____. *Modelo de prueba de lenguaje y comunicación*. Santiago: DEMRE, 2018c. Disponible en: <<https://psu.demre.cl/publicaciones/pdf/2019-18-07-19-modelo-lenguaje.pdf>>. Consultado el: 22 mar. 2019.
- ELLIOT, N. *On a scale: a social history of writing assessment in America*. New York: Peter Lang, 2005.
- EODICE, M.; GELLER, A. E.; LERNER, N. *The meaningful writing project: learning, teaching, and writing in higher education*. Boulder, CO: Utah State University Press, 2016.
- FLOTTS, M. P. et al. *Aportes para la enseñanza de la escritura*. Santiago: UNESCO, 2016.
- GARDNER, S.; NESI, H. A classification of genre families in university student writing. *Applied Linguistics*, Oxford, v. 34, n. 1, p. 25-52, 2013.
- GÓMEZ VERA, G. et al. *La producción escrita de estudiantes y escuela en base a resultados SIMCE, factores contextuales y modelos de buenas prácticas docentes*. Chile: FONIDE, 2016. Disponible en: <<https://centroestudios.mineduc.cl/wp-content/uploads/sites/100/2017/07/INFORME-FINAL-F911437.pdf>>. Consultado el: 22 oct. 2018.
- HUOT, B. Toward a new theory of writing assessment. In: HUOT, B.; O'Neill, P. (Ed.). *Assessing writing: a critical sourcebook*. Boston: Bedford & St. Martin's, 2009. p. 160-175.
- KLIEME, E.; HARTIG, J.; RAUCH, D. The concept of competence in educational contexts. In: J. HARTIG, J.; KLIEME, E.; LEUTNER, D. (Ed.). *Assessment of competencies in educational contexts*. Ashland, OH, US: Hogrefe & Huber Publishers, 2008. p. 3-22.
- LEMAITRE, M. J. Aseguramiento de la calidad: una política y sus circunstancias. In: Bernasconi, A. (Ed.). *La educación superior de Chile: transformación, desarrollo y crisis*. Santiago: Ediciones UC, 2015. p. 295-344.

- LERNER, D. *Leer y escribir en la escuela: lo real, lo posible y lo necesario*. México: Fondo de Cultura Económica, 2001. p. 25-37.
- LUCE-KAPLER, R.; KLINGER, D. Uneasy writing: the defining moments of high-stakes literacy testing. *Assessing Writing*, [S.l.], v. 10, p. 157-173, 2005.
- MATEOS, M.; MARTÍN, E.; VILLALÓN, R. La percepción de profesores y alumnos en la educación secundaria sobre las tareas de lectura y escritura que realizan para aprender. In: POZO, J. I. et al. (Ed.). *Nuevas formas de pensar la enseñanza y el aprendizaje*. Barcelona: Graó, 2006. p. 231-242.
- MISLEVY, R. J. *Sociocognitive foundations of educational measurement*. London: Routledge, 2018.
- MURPHY, S.; YANCEY, K. B. Construct and consequence: validity in writing assessment. In: BAZERMAN, C. (Ed.). *Handbook of research on writing: history, society, school, individual, text*. New York: Taylor & Francis, 2008. p. 448-473.
- NAVARRO, F. Más allá de la alfabetización académica: las funciones de la escritura en educación superior. In: ALVES, M. A.; BORTOLUZZI, Valéria Iansen (Ed.). *Formação de professores: ensino, linguagens e tecnologias*. Porto Alegre, RS: Editora Fi, 2018. p. 13-49.
- _____. F. Aportes para una didáctica de la escritura académica basada en géneros discursivos. *DELTA: documentação de estudos em linguística teórica e aplicada*, São Paulo, v. 35, n. 1, p. 1-32, 2019.
- OLSON, D. R. *El mundo sobre el papel: el impacto de la escritura y la lectura en la estructura del conocimiento*. Barcelona: Gedisa, 1998.
- O'NEILL, P.; MOORE, C.; HUOT, B. *A guide to college writing assessment*. Logan, Utah: Utah State University Press, 2009.
- _____. How does Writing assessment frame college writing programs?. In: _____; PERELMAN, L. (Ed.). *Writing assessment in the 21st century: essays in honor of Edward M. White*. New York: Hampton Press, 2012. p. 439-456.
- PEARSON. *Final report: evaluation of the Chile PSU*. Santiago: Pearson, 2013. Disponible en: <http://www.educacion2020.cl/sites/default/files/2013013111057540.chile_psu-finalreport.pdf>. Consultado el: 22 oct. 2018.
- PELLEGRINO, J. W.; HILTON, M. L. (Ed.). *Education for life and work: developing transferable knowledge and skills in the 21st century*. Washington, DC: The National Academies Press, 2012.
- PERELMAN, L. *Towards a new NAPLAN: testing to the teaching*. Surry Hills: NSW Teachers Federation, 2018.
- POE, M.; INOUE, A.; ELLIOT, N. (Ed.). *Writing assessment, social justice, and the advancement of opportunity*. Fort Collins, Colorado: The WAC Clearinghouse, 2018.
- PRIOR, P.; LUNSFORD, K. History of Reflection and Research on Writing. In: BAZERMAN, C. (Ed.). *Handbook of research on writing: history, society, school, individual, text*. New York: Taylor & Francis, 2008. p. 81-96.

- READ, J. *Assessing vocabulary*. Cambridge: Cambridge University Press, 2000.
- ROSE, D.; MARTIN, J. R. *Learning to write, reading to learn: genre, knowledge and pedagogy in the Sydney school*. London: Equinox, 2012.
- ROSLI, N.; CARLINO, P.; RONI, C. Retención escolar y educación de calidad: logros y desafíos pendientes en una escuela secundaria Argentina. *Archivos Analíticos de Políticas Educativas*, Arizona, v. 23, n. 102, p. 2-27, 2015.
- SANTELICES, M. V.; GALLEGUILLOS, P.; CATALÁN AVENDAÑO, X. El acceso y la transición a la universidad en Chile. In: BERNASCONI, A. (Ed.). *La educación superior de Chile: transformación, desarrollo y crisis*. Santiago: Ediciones UC, 2015. p. 581-627.
- SCHLEPPEGRELL, M. J. *The language of schooling: a functional linguistics perspective*. Mahwah, NJ: Lawrence Erlbaum, 2004.
- SCHMANDT-BESSERAT, D.; ERARD, M. Origins and forms of writing. In: BAZERMAN, C. (Ed.). *Handbook of research on writing: history, society, school, individual, text*. New York: Taylor & Francis, 2008. p. 7-26.
- SLOMP, D. H. Harming not helping: The impact of a Canadian standardized writing assessment on curriculum and pedagogy. *Assessing Writing*, [s.l.], v. 13, p. 180-200, 2008.
- STARKE-MEYERRING, D.; PARÉ, A. The roles of writing in knowledge societies: questions, exigencies, and implications for the study and teaching of writing. In: STARKE-MEYERRING, D. et al. (Ed.). *Writing in knowledge societies*. Colorado: The WAC Clearinghouse; South Carolina: Parlor Press, 2011. p. 3-28.
- VERA, J. *El sistema de admisión a la universidad: permanencia y cambio*. Santiago: Editorial Universitaria, 2015.
- WESTERA, W. Competences in education: a confusion of tongues. *Journal of Curriculum Studies*, Basingstoke, v. 33, n. 1, p. 75-88, 2001.
- WHITE, E. M. Holisticism. *College Composition and Communication, National Council of Teachers of English*, [s.l.], v. 35, n. 4, p. 400-409, 1984.
- WIGGINS, G. The case for authentic assessment: practical assessment. *Research & Evaluation*, Guildford, v. 2, n. 2, p. 1-3, 1990. Disponible en: <<http://PAREonline.net/getvn.asp?v=2&n=2>>. Consultado el: 22 oct. 2018.
- WILLIAMSON, M. An introduction to holistic scoring. In: _____; HUOT, B. (Ed.). *Validating holistic scoring for writing assessment: theoretical and empirical foundations*. Cresskill, NJ: Hampton Press, 1993. p. 206-232.
- ZAVALA, V. La escritura académica y la agencia de los sujetos. *Cuadernos Comillas*, [s.l.], v. 1, p. 52-66, 2011.

Recebido em: 18/11/2018

Aceito para publicação em: 01/03/2019

Validity and fairness: Towards a meaningful assessment in standardized writing tests

Abstract

Until recently, validity of standardized writing tests was mostly based on interrater reliability, writing tasks were not informed by theory and research on literacy, texts were assessed using proxy measures or analytical rubrics that compartmentalized writing traits, and, most importantly, contextual validation, fairness and consequences of uses of tests' results were not considered. This paper presents the development of a new writing assessment to improve the information on student's written performance available at the beginning of tertiary education in Chile. The new test includes prompts to write for academic purposes and social participation; provides details on the rhetorical situation, purposes, audiences, and assessment expectations; gives room for planning and revision; and rates texts according to a four-point holistic rubric that focuses on purpose achievement. Results show high levels of students' engagement and normal distribution of performance.

Keywords: Writing. Standardized assessment. Higher Education. Discursive and textual genres. Validity.

Validade e justiça: por uma avaliação significativa em provas estandarizadas de escrita

Resumo

Até pouco tempo, a validade de provas estandarizadas de escrita era baseada na confiabilidade interavaliadores, as tarefas não estavam informadas por teoria e investigação, os textos eram avaliados por meio de indicadores indiretos ou rubricas analíticas que fragmentavam seus componentes e, sobretudo, a validação contextual, a justiça e as consequências do uso dos resultados não eram consideradas. O presente artigo apresenta um novo teste de avaliação de escritura desenvolvido para fornecer informações sobre o desempenho escrito no início do Ensino Superior no Chile. Inclui tarefas para escrever com fins acadêmicos e com participação cidadã; detalha a situação, os propósitos, as audiências e as expectativas; habilita o planejamento e a revisão; e avalia os textos através de uma rubrica holística centrada no cumprimento de objetivos. Os resultados de sua aplicação mostram um alto nível de interesse dos estudantes e uma distribuição normal dos desempenhos.

Palavras-chave: Escrita. Avaliação estandarizada. Educação superior. Gêneros discursivos. Validação.