

## La medición del crecimiento académico contextualiza la complejidad del texto

### Measuring Academic Growth Contextualizes Text Complexity

Gary L. Williamson

MetaMetrics®, North Carolina, EE.UU.

#### Resumen

Para una medición óptima del crecimiento académico se requiere una escala unidimensional, continua, a iguales intervalos, de desarrollo e invariable con respecto a la ubicación y el tamaño de la unidad. Estas características pueden obtenerse combinando la medición Rasch con una teoría de constructos de lectura operacionalizada y el anclaje de la escala resultante en dos puntos en un continuo de complejidad de texto. De esta manera, personas y textos se juntan en una escala de desarrollo común. Así, una iniciativa de política educativa de Estados Unidos de América (EE.UU.) recomendó incrementar la exposición de los estudiantes a textos complejos. El estándar de exposición a textos puede examinarse en forma productiva a la luz del crecimiento histórico de la lectura si ambos se miden con una escala común de base teórica. El modelado paramétrico de curvas alternativas de crecimiento promueve un debate sobre cómo los estudiantes pueden alcanzar el objetivo de estar preparados para la universidad y para una carrera.

**Palabras clave:** crecimiento, lectura, complejidad de texto, medición

---

#### Correspondencia a:

Gary L. Williamson  
MetaMetrics®, North Carolina, USA.  
1000 Park Forty Plaza Drive, Suite 120  
Durham, NC 27713, USA.  
Correo electrónico: gwilliamson@lexile.com

---

© 2015 PEL, <http://www.pensamientoeducativo.org> - <http://www.pel.cl>

ISSN: 0719-0409      DDI: 203.262, Santiago, Chile  
doi: 10.7764/PEL.52.2.2015.6

---

## Abstract

---

Optimally measuring academic growth requires a scale that is unidimensional, continuous, equal-interval, developmental, and invariant with respect to location and unit size. These characteristics can be attained by combining Rasch measurement with an operationalized reading construct theory and anchoring the resulting scale at two points on a text-complexity continuum. Hence, persons and texts are brought onto a common developmental scale. Coincidentally, an educational policy initiative in the United States of America (USA) recommended increasing students' exposure to complex texts. The text-exposure standard can be productively examined in light of historical reading growth when both are measured with a common theory-based scale. Parametric modeling of alternative growth curves motivates a conversation about how students may attain a college and career readiness goal.

*Keywords:* growth, reading, text complexity, measurement

En la primavera de 2014, realicé una presentación para la sección de Metrología y Resultados del 17° Taller Internacional de Medición Objetiva [International Objective Measurement Workshop (IOMW)]. Durante esa sesión, describí varios ejes de trabajo relacionados entre sí que, en mi opinión, reflejan el poder de la medición para informar dos temas de reciente aparición en la práctica educacional: el crecimiento y la complejidad del texto. En este artículo, resumo esas ideas y describo ejemplos de apoyo que ilustran el potencial práctico de la aplicación de la medición Rasch a problemas educacionales. Sin embargo, puede resultar útil que primero entregue algo de contexto para motivar la investigación y para definir algunos términos que podrían ser poco conocidos para el lector principiante.

## Contextualización para el lector

### Investigación en lectura y pedagogía

En un artículo retrospectivo sobre investigación en lectura y pedagogía de la lectura en los Estados Unidos de América (EE.UU.), Pearson (2004) sostuvo que la enseñanza de la lectura y la investigación en lectura habían estado sujetas durante décadas a repetidos «movimientos pendulares» que se iban alternando entre prácticas pedagógicas (whole language vs. instrucción basada en habilidades), junto con sus correspondientes paradigmas de investigación (experimentos aleatorizados vs. investigaciones cualitativas y naturalistas). Más aún, indicó que la instrucción lectora y la investigación en lectura habían sido moldeadas por fuerzas políticas opuestas, las que favorecían a uno u otro enfoque. Pearson llamó a este escenario «Las guerras de la lectura» [The Reading Wars]. En su artículo, Pearson hacía un llamado a buscar mayor equilibrio, y observaba: «Es nuestra obligación moral y ética usar la mejor evidencia que podamos obtener para tomar decisiones de alto impacto con respecto a las políticas educacionales» (p. 21).

Durante la última década, probablemente parte del equilibrio recomendado por Pearson (2004) se ha aplicado a textos de lectura centrales (core reading texts) y a énfasis educativos empleados en los primeros años de la enseñanza escolar (Comunicación personal, J. Fitzgerald, 25 de marzo de 2015). Sin embargo, los puntos de vista sobre la lectura siguen bifurcándose de acuerdo a ciertas líneas: los polos hoy en día parecen corresponder a los enfoques cognitivo (por ej., Kendeou, van den Broek, Helder, & Karlsson, 2014) y al sociocultural (por ej., Alvermann, Hinchman, Moore, Phelps, & Waff, 2006). Los lectores pueden encontrar en Alvermann, Unrau y Ruddell (2013) una perspectiva de los dos extremos del espectro. En consecuencia, la medición del logro en lectura sigue estando guiada por diferentes visiones de qué es la lectura, cuál es la mejor forma de enseñarla y cómo cuantificar el logro en lectura (si es que debiera hacerse).

### Acciones relevantes en política educacional en EE.UU.

Durante la primera mitad del siglo XX, no existía una manera definitiva de determinar la habilidad lectora media de los estudiantes en EE.UU. porque no había una medida educativa de logro en lectura de uso común. Existían muchos tests de lectura, pero entregaban resultados no comparables. Para paliar la situación, se creó una Evaluación Nacional de Progreso Educativo [National Assessment of Educational Progress (NAEP)] durante la década de 1960 para recoger datos sobre desempeño educativo. El nuevo programa de evaluación se administró por primera vez en 1969, y los primeros datos sobre logro lector se obtuvieron durante el año escolar 1970–71. En la actualidad, la NAEP sigue recolectando datos de logro lector de forma periódica. Sin embargo, la recolección de datos de la NAEP está diseñada para producir resúmenes globales nacionales o estatales sobre el desempeño lector de los estudiantes, sin entregar puntajes individuales de cada alumno.

En EE.UU., las decisiones educativas se toman predominantemente a nivel estatal y de las comunidades locales. El rol históricamente limitado del gobierno federal en la educación se incrementó hasta cierto punto cuando el Congreso de EE.UU. aprobó la Ley de Educación Primaria y Secundaria de 1965 [Elementary and Secondary Education Act (ESEA)]. Desde el comienzo, el objetivo central de la ESEA fue mejorar las oportunidades educativas de todos los estudiantes. La ley ha sido periódicamente actualizada o revisada. La más reciente revisión de la ESEA consistió en la Ley *Que ningún niño quede atrás* (No Child Left Behind Act) de 2001 (No Child Left Behind [NCLB], 2002), que entró en vigor con la firma del Presidente George W. Bush en enero de 2002. Desde ese momento, la ESEA se ha denominado NCLB. La Ley NCLB exige que todas las escuelas públicas midan el logro de sus estudiantes en lectura y matemáticas durante grados específicos. La exigencia federal para medir e informar los resultados del logro en lectura individual de los alumnos constituyó un enorme cambio en las prácticas de medición educacional en EE.UU.

Sin embargo, el requerimiento federal para medir el logro de los estudiantes en lectura no especificaba qué prueba(s) debían usar los estados. Los estados conservaron la autoridad para crear sus propios currículos de lectura estatales y para seleccionar o generar sus propias mediciones de logro en lectura. Como resultado de esta situación, en el conjunto de los estados se utiliza una multitud de mediciones de lectura; asimismo, cada estado establece sus propios estándares sobre desempeño estudiantil aceptable, de acuerdo a la medición de logro en lectura que haya adoptado. Esto ha ido generando una disonancia cognitiva muy extendida, ya que los resultados de desempeño estudiantil derivados de evaluaciones estatales no son comparables entre sí ni tampoco con los resultados de la NAEP.

En EE.UU., cada estado tiene un director ejecutivo, llamado gobernador, que es elegido popularmente. Desde 1908, la Asociación Nacional de Gobernadores [National Governors Association (NGA)] ha funcionado como una organización bipartidista que apoya varias actividades de liderazgo para los gobernadores de la nación. En particular, una de sus iniciativas es el Centro para las Buenas Prácticas (Center for Best Practices), que promueve la identificación y difusión de información para mejorar las prácticas educativas en todo el país. Los miembros de la NGA son los gobernadores de los estados, territorios y estados libres asociados (Commonwealths) de EE.UU.

Cada estado posee también un jefe de las escuelas estatales, quien está a cargo de las escuelas públicas. En algunos estados, el jefe de las escuelas estatales es elegido popularmente, mientras que en otros puede ser designado ya sea por el gobernador o por uno o más delegados de este. El Consejo de Jefes de Escuelas Estatales [Council of Chief State School Officers (CCSSO)] es un organismo nacional compuesto por los líderes educacionales de cada estado.

Luego de la aprobación de la ley NCLB, la rendición de cuentas con respecto al desempeño de las escuelas se convirtió en un tema central en el ámbito educativo. La falta de comparabilidad entre los resultados estatales y la aparente desconexión entre los informes estatales de logro y la NEAP fueron causando cada vez más preocupación entre los gobernadores y los jefes de escuelas estatales en EE.UU. En respuesta al problema, estas dos organizaciones comenzaron a trabajar en conjunto para mejorar las prácticas educativas y la rendición de cuentas del sistema escolar.

La CCSO y el Centro para las Buenas Prácticas de la NGA se enfocaron en involucrar a los educadores en el desarrollo de un marco común de metas consistentes de aprendizaje escolar dirigidas a que los estudiantes estuvieran preparados para la educación superior y el mundo laboral al momento de terminar la educación secundaria. Luego de un extenso período de desarrollo, sugerencias y modificaciones, la NGA y el CCSSO publicaron los Estándares Comunes Básicos Estatales [Common Core State Standards (CCSS)] en junio de 2010. Cada estado revisó individualmente los estándares y, para fines de 2013, los CCSS se habían adoptado ampliamente. En junio de 2014, 43 estados, el Distrito de Columbia, cuatro territorios y el Departamento de Actividades de Educación en Defensa [Department of Defense Education Activity (DoDEA)] habían adoptado los CCSS y se encuentran implementando los estándares de acuerdo a su propia calendarización.

Durante el proceso de implementación de los CCSS, los estados comenzaron a notar que necesitaban prestar atención a la evaluación del desempeño de los estudiantes en relación con los estándares. Dada esta situación, emergieron dos consorcios de estados con filosofías de evaluación ligeramente distintas: el Consorcio de Evaluación más Inteligente y Equilibrada [Smarter Balanced Assessment Consortium (SBAC)] y la Asociación para la Evaluación de la Preparación para la Educación Superior y el Mundo Laboral [Partnership for Assessment of Readiness for College and Careers (PARCC)]. Aunque sus principios orientadores no son los mismos, ambos consorcios buscan mejorar las prácticas de evaluación y aumentar la comparabilidad de los resultados de desempeño estudiantil.

Incluso antes de las llamadas «guerras de la lectura», una característica importante de la educación para la lectura y la investigación en lectura en EE.UU. era la falta de un acuerdo sobre qué es la lectura; lógicamente, esto llevó a la aparición de múltiples instrumentos de medición de la lectura disponibles para su uso en educación. Aunque los CCSS representan un acuerdo común ampliamente aceptado acerca de las metas de aprendizaje de los estudiantes y, a pesar de que el SBAC y la PARCC han hecho avances en el diseño de pruebas alineadas con los CCSS, aún no existe una medición común de la comprensión lectora individual de los estudiantes en EE.UU. ni un acuerdo común sobre el nivel de logro en lectura que estos deben alcanzar.

### Contexto de medición

Realizar una revisión general del campo de la medición educacional supera con creces los límites del presente artículo. Para acceder al estudio enciclopédico más reciente sobre este tema, recomiendo que los lectores interesados consulten Brennan (2006), un volumen de 779 páginas patrocinado por el Consejo Nacional de Medición en Educación [National Council on Measurement in Education] y el Consejo de Educación de EE.UU. [American Council on Education]. En lugar de ello, me referiré a tres aspectos de la medición educacional (objetividad específica, objetividad general y medición conjunta) relativos a los ejemplos que se presentan más adelante en este artículo. Por lo tanto, me refiero –solo de manera general– a los avances que dieron lugar a estas tres ideas.

La medición educacional emergió en EE.UU. a comienzos del siglo XX. A mediados de siglo, el conocimiento existente sobre la teoría de la evaluación estaba siendo canonizado como *Teoría Clásica de los Tests* (por ej., Gullicksen, 1950). Este enfoque continuó enseñándose de modo primario, a pesar de que los nuevos modelos de respuesta al ítem [*ítem response theory* (IRT)] (por ej., Lord, 1952; Lord & Novick, 1968; Rasch, 1960) ya comenzaban a suplir y expandir el campo de la psicometría. Gradualmente, la teoría de respuesta al ítem comenzó a usarse ampliamente (Lord, 1980) y hoy en día es considerada esencial para desarrollar pruebas educacionales. Su ventaja central sobre la teoría clásica de test es que la IRT entrega un modelo matemático explícito para caracterizar la respuesta de un individuo a un ítem en una prueba.

Matemáticamente, el modelo más simple de IRT es el que se le atribuye a Rasch (1960). Rasch (1966, 1977) fue notando de forma gradual que su formulación matemática tenía propiedades interesantes, las que consideró epistemológicamente importantes. Por ejemplo, al calcular la proporción entre las posibilidades de respuesta correcta (a un ítem) de dos personas, los parámetros del ítem se cancelan en la expresión. «Entonces, el modelo hace aseveraciones *objetivas o separadas del ítem* sobre la probabilidad relativa de que dos personas respondan correctamente a un ítem o a un conjunto de ítems, sin referencia a los ítems mismos» (Thissen & Wainer, 2001, p. 75, cursiva en el original). Otra

consecuencia directa del modelo es que al calcular los odds ratio de que una persona responda correctamente a un ítem, arroja la diferencia aritmética entre la capacidad de la persona y la dificultad del ítem. Bond y Fox (2001) sugirieron que la capacidad de comparar personas e ítems directamente, sin referencia a otros ítems o personas, produce el tipo de medición «que esperamos en las ciencias físicas» (p. 203). La capacidad de producir calibraciones de ítems no relativas a personas y mediciones de personas sin referencia a ítems es propia del modelo Rasch. Esta propiedad matemática del modelo Rasch se ha denominado *separación de parámetros*; Rasch (1977) la llamó *objetividad específica*.

La IRT entregó un nuevo marco para el desarrollo de tests educacionales al modelar qué ocurre cuando los individuos responden a los ítems, considerando las características de estos ítems. El modelo Rasch va más allá al asegurarse de que las escalas de medición no dependan de los individuos o ítems específicos empleados en su construcción. Sin embargo, a pesar de que los modelos IRT proponen un constructo de persona (por ej., habilidad del lector) y una dimensión de tarea (por ej., complejidad del texto), estos modelos en sí mismos no entregan detalles explícitos sobre qué son estos constructos. Más aún, los modelos IRT no emplean un marco de referencia absoluto para calibrar sus escalas. Por lo tanto, una prueba de lectura construida con IRT no es directamente comparable con otra prueba de lectura que también haya sido construida con IRT.

¿Es posible crear una escala absoluta de habilidad lectora que pueda ser interpretada universalmente en términos del constructo mismo? ¿Cómo avanzar desde la objetividad específica a la objetividad general en la medición educacional? Se requeriría de una definición operacional clara del continuo de constructos y el compromiso de emplear un marco de referencia consistente para la escala de medición de constructos.

En la década de 1980 se logró un avance enorme cuando dos empresas de medición intentaron, de forma independiente, construir mediciones generalmente objetivas de habilidad lectora, combinando el modelo de medición Rasch con una especificación operacional explícita del continuo de tareas lectoras. Como es el caso de otros modelos IRT, el modelo Rasch permite ubicar tanto a las personas como a las tareas (ítems) en una escala común, pero la ubicación de la escala es indeterminada. La innovación clave involucró dos pasos que anclaban la escala y definían su tamaño de unidad en términos de un continuo de tareas lectoras bien especificado.

En primer lugar, los investigadores cuantificaron la dificultad lectora de los textos en términos de sus características semánticas y sintácticas, las que representan adecuadamente la demanda cognitiva experimentada por los individuos al leer. Esencialmente, este paso crea una especificación (o modelo) cuantitativa de constructos para el continuo de tareas de lectura. En segundo lugar, los investigadores demostraron que las dificultades empíricas de tipos de ítems específicos, bien definidos y basados en texto podían predecirse casi a la perfección sobre la base de la complejidad de los textos asociados a los ítems. Una vez que los estudiantes y los ítems fueron medidos a través del modelo Rasch, la especificación de la teoría del constructo se usó para calibrar los ítems con respecto al continuo de dificultad textual. Esto generó una correspondencia directa entre las mediciones de persona y de texto en términos de un continuo existente en el mundo real y basado en la complejidad del texto.

La empresa hoy conocida como Questar Assessment, Inc. fue la primera en emplear este tipo de enfoque. Esta desarrolló el programa Grados de Poder de Lectura [Degrees of Reading Power® (DRP®)], que reporta mediciones lectoras de estudiantes en pruebas referidas a un criterio dentro de una Escala de Complejidad Textual DRP (DRP Scale of Text Complexity), de tipo propietario, la cual es empleada para medir la dificultad de leer material impreso (Questar Assessment, Inc., 2012). Nelson, Perfetti, D. Liben y M. Liben (2012) describieron la escala de la siguiente forma:

La dificultad DRP de un texto se expresa en unidades DRP en una escala continua que posee un rango teórico de 0 a 100. En la práctica, el texto común en inglés va desde 25 a 85 unidades DRP, representando los valores más altos una mayor dificultad del texto (p. 11).

Questar definió un «modelo de comprensión de prosa» DRP basado en la aplicación de la fórmula de legibilidad Bormuth (1969) para medir complejidad textual. Su tipo de ítem de referencia era un ejercicio de completación inserto en un texto, administrado de acuerdo a un protocolo específico. El tamaño de unidad de la escala DRP se especificó en términos de una transformación de la medida

Bormuth de complejidad textual, R. La investigación ha mostrado que la escala DRP ubica tanto la capacidad de lectura del estudiante como la complejidad del texto en una escala común, bien definida y unidimensional, que se mantiene invariante a través del tiempo. De esta manera, la investigación apoya la idea de que las pruebas DRP «son como las mediciones en las ciencias naturales» (Koslin, Zeno, & Koslin, 1987, p. 171).

Casi simultáneamente, una segunda compañía siguió la misma idea fundamental. MetaMetrics® desarrolló el Marco Lexile® para la Lectura (Lexile® Framework for Reading) para medir a los lectores y a los textos en una escala común. El Marco Lexile, en lo fundamental, también se apoya en una aplicación del modelo Rasch. Sin embargo, nuevamente, el modelo Rasch fue expandido para que alcanzara objetividad general, cuantificando explícitamente el continuo de tareas. El proceso incluyó tres pasos. Primero, se diseñó un ítem de referencia específico y basado en texto para medir la habilidad del lector. Para responder al ítem, los individuos debían leer y comprender un segmento de un texto en prosa. Segundo, se desarrolló una *ecuación de especificación de constructo* (Stenner, Smith, & Burdick, 1983) para cuantificar la complejidad textual de un material de lectura en prosa sobre la base de las características del material. Tercero, se demostró empíricamente que las mediciones de complejidad textual producidas por la ecuación de especificación podían predecir, con gran precisión, las dificultades relativas al modelo Rasch de los ítems de referencia (Burdick & Stenner, 1996). En consecuencia, se permitió expresar la dificultad del ítem y la habilidad de la persona mediante el modelo Rasch en una escala invariante común definida por el continuo de complejidad textual.

Con objeto de definir una unidad de escala lógica para la escala Lexile, MetaMetrics decidió anclar explícitamente su escala en dos puntos dentro del continuo de complejidad textual. Sobre la base de este anclaje, una unidad de la escala Lexile es igual a 1/1000 de la diferencia entre la legibilidad de ciertos libros básicos para niños y la legibilidad de una enciclopedia en línea para adultos (Stenner, H. Burdick, Sanford, & D. S. Burdick, 2007). Debe tenerse en cuenta que anclar la escala en dos puntos bien definidos es suficiente para definir el tamaño de unidad de la escala sin ambigüedad. Este enfoque produjo una unidad de medida bien definida que mantiene su tamaño absoluto en distintas mediciones. Se puede notar que este método es directamente análogo a la forma en que se estandarizó el metro basándose en la longitud del arco de meridiano que pasa por París (Legendre, 1805). También es exactamente análogo a la manera en que se anclan las escalas de temperatura. Irónicamente, algunos años después, esta técnica fue recomendada por el Presidente saliente del Consejo Nacional de Medición en Educación (National Council for Measurement in Education) como una característica deseable para incluir en las escalas de logro académico; asimismo, el directivo desafió a la comunidad del campo de la medición a desarrollar escalas de este tipo para su uso en educación (Reckase, 2009), aparentemente sin saber que esto ya se había hecho para la habilidad lectora.

El modelo Rasch expresa la probabilidad de responder correctamente a un ítem como la diferencia exponencial entre la habilidad de una persona y la dificultad del ítem. Anteriormente se mencionó que, debido a la separación de parámetros, el logaritmo del odds de una respuesta correcta es la diferencia aritmética entre la habilidad de la persona y la dificultad del ítem. Estos hechos son manifestaciones de otra propiedad importante del modelo de medición Rasch: que produce medición conjunta. La medición conjunta permite poner al mismo tiempo dentro de una escala dos variables que predicen un resultado conjuntamente. Por ejemplo, la habilidad del lector y la dificultad del texto predicen conjuntamente la comprensión lectora; luego, tanto la medición del lector como la medición de la dificultad del texto pueden ubicarse en una escala común a través de una medición conjunta.

A través de la medición conjunta con el modelo Rasch, tanto el Marco Lexile y la escala DRP pueden utilizarse para generar puntajes de habilidad lectora de los estudiantes que se reportan dentro en un continuo de complejidad textual, lo que le suplementa la interpretabilidad de estos puntajes en términos de un continuo de tareas de lectura del mundo real. El uso primario de estos dos sistemas ha sido emparejar a los estudiantes con textos de una dificultad apropiada. Los ejemplos presentados más adelante en el artículo ilustran que la metodología de anclar las mediciones de lectura de los estudiantes a un continuo bien definido de complejidad textual tiene mucho potencial para medir el crecimiento de los estudiantes en lectura.

## Tecnologías de medición utilizadas para los ejemplos

Los ejemplos presentados en este artículo se basan en múltiples elementos de trabajos previos llevados a cabo por un gran número de individuos y organizaciones. Una categoría de trabajo consistió en medir complejidad textual en colecciones de textos, lo que produjo distribuciones de mediciones de complejidad textual para materiales de lectura característicos de las escuelas públicas y de los ámbitos post-secundarios en EE.UU. Otra categoría de trabajo produjo mediciones longitudinales de logro en lectura para individuos que asistían a escuelas públicas en el estado de Carolina del Norte. La tercera categoría de trabajo consistió en modelos de crecimiento multinivel aplicados a los datos longitudinales.

Hace aproximadamente 20 años, el Departamento de Instrucción Pública de Carolina del Norte decidió usar el Marco Lexile para la Lectura como escala suplementaria en sus Pruebas de Comprensión Lectora de Fin de Grado [End-of-Grade Tests of Reading Comprehension] de 3ro a 8vo grado, de modo de entregarle mayor utilidad instruccional a los resultados de sus pruebas (específicamente, ayudando a los profesores a emplear textos de dificultad apropiada para sus estudiantes). Los ejemplos presentados en este artículo ilustran la utilidad adicional de la escala para modelar el crecimiento lector de los alumnos y para realizar interpretaciones conjuntas de este en relación con requerimientos de complejidad textual. En consecuencia, el origen de estas fuentes de datos es de interés contextual y se delinea brevemente a continuación.

**Medición textual.** El Marco Lexile para la Lectura basa sus mediciones de complejidad textual en las características semánticas y sintácticas de los textos. Específicamente, la complejidad textual se cuantifica como una combinación lineal de dos variables: la frecuencia de las palabras y la longitud de las oraciones (Stenner, H. Burdick, Sanford, & D. S. Burdick, 2007). Las mediciones Lexile de complejidad textual se calculan para texto en prosa editado profesionalmente. Los textos que no tienen una puntuación convencional (por ej., poemas, letras de canciones o recetas) no se miden con este marco.

Los textos deben prepararse para ser medidos de acuerdo a convenciones específicas. Por ejemplo, antes de analizar un texto, se eliminan todos los encabezados y nombres de sección (porque no son oraciones completas). Los textos son digitalizados para facilitar el análisis. La producción de mediciones textuales Lexile se realiza de forma automática con un programa llamado Lexile® Analyzer, que está disponible sin costo para los educadores o bien puede usarse profesionalmente adquiriendo una licencia.

A lo largo de los años, se han llevado a cabo numerosos estudios para medir la dificultad de lectura o la complejidad textual de los libros de texto usados en las escuelas públicas de EE.UU. Está más allá de los objetivos de este artículo realizar una exhaustiva revisión o crítica de esa literatura. En lugar de ello, describiré brevemente tres publicaciones recientes que entregan antecedentes relevantes sobre las colecciones de textos específicas usadas en los ejemplos presentados en este artículo.

Primero, Williamson, Koons, Sandvik y Sanford-Moore (2012) describieron un continuo de complejidad textual para los grados del 1 al 12. Los autores identificaron textos consultando listas de adopción de libros de texto en siete estados (Florida, Georgia, Indiana, Carolina del Norte, Oregon, Texas y Virginia), donde era común que se adoptaran libros de texto a nivel estatal. Se centraron en ediciones de libros de texto para estudiantes, agrupadas en seis categorías de contenido (Salud, Lenguaje, Literatura, Matemáticas, Ciencia y Estudios Sociales), para aportar una cobertura curricular diversa. Los libros de texto que aparecieron en las listas de adopción de varios estados se seleccionaron para el estudio, suponiendo que estos probablemente serían usados por un gran número de estudiantes. Si un libro de texto pertenecía a una serie, todos los libros de la serie se incluían en el estudio. Los textos diseñados para un solo grado fueron seleccionados para los análisis de modo de caracterizar inequívocamente las distribuciones de complejidad textual intra-nivel. En total, el estudio incluyó 487 libros de texto desde los grados 1 al 12. Los autores observaron que la complejidad media de los textos aumentaba de acuerdo al nivel, casi monótonamente, y que la complejidad variaba dentro de cada nivel.

Williamson (2008) consideró materiales de lectura presentes en varios contextos de la vida post-secundaria (por ej., las universidades, los institutos profesionales, los lugares de trabajo, el ejército y la vida cívica). Este investigador presentó un continuo de dificultad textual para los dominios de actividad

post-secundarios y cuantificó las brechas de complejidad textual entre los textos de las escuelas secundarias (empleados en los grados 11 y 12) y los usados en el mundo post-secundario. Notablemente, halló una brecha de 265L entre las medianas de los textos de las escuelas secundarias y de los textos universitarios.

Stenner, Sanford-Moore y Williamson (2012) extendieron el conocimiento de la comunidad científica acerca de los materiales de lectura post-secundarios. Su trabajo expandió el aporte de Williamson (2008) al aumentar las colecciones de textos de las instituciones educacionales post-secundarias y al incluir dos categorías adicionales de textos. Enfocándose en tres estados del sur de EE.UU. (Georgia, Tennessee y Texas), añadieron textos adicionales correspondientes a universidades y a institutos de estudios superiores; además, incluyeron textos de centros de formación técnica. Asimismo, los autores incorporaron muestras de textos tomadas de diarios internacionales en inglés y artículos destacados seleccionados de Wikipedia. Combinando todas las fuentes de textos post-secundarios ( $n = 2990$ ), reportaron que la complejidad media de estos llegaba a 1300L, con un rango intercuartílico entre 1200L y 1380L.

**Mediciones de habilidad lectora de los estudiantes.** Las medidas Lexile de habilidad lectora derivan de evaluaciones de lectura que se han enlazado psicométricamente a la escala Lexile. Dada una prueba objetivo, la cual se enlazará, el proceso involucra la creación de una Prueba de Enlace Lexile (PEL) [Lexile Linking Test (LLT)], rigurosamente diseñada para: (a) medir el mismo constructo que la prueba objetivo, (b) tener la misma confiabilidad que la prueba objetivo y (c) tener las mismas especificaciones que la prueba objetivo en todos los aspectos excepto uno: la PEL solo incluye el tipo de ítem estándar asociado con el Marco Lexile. Tanto la PEL como la prueba objetivo se aplican a una muestra de estudiantes, luego de lo cual se deriva una función de enlace simétrica. De este modo, los puntajes de la prueba objetivo se expresan en la escala Lexile. Los estudios de enlace Lexile cumplen al menos tres de los cinco requisitos especificados por Holland y Dorans (2006) para considerar que alcanzan la equiparación, el tipo de enlace psicométrico más fuerte.

Algunos de los ejemplos incluidos en este artículo ilustran el crecimiento en lectura y usan los resultados de investigaciones anteriores descritas por Williamson (2014). Sus datos sobre lectura derivaban de aplicaciones de las Pruebas de Comprensión Lectora de Fin de Grado de Carolina del Norte (North Carolina End-of-Grade [NC EOG] Reading Comprehension Test) (Sanford, 1996), que el estado enlazó con el Marco Lexile para la Lectura. Las características de los datos de los estudiantes se resumen brevemente aquí para conveniencia del lector interesado.

Las Pruebas de Comprensión Lectora de Fin de Grado de Carolina del Norte (1ra Edición), eran pruebas de opción múltiple aplicadas en formato de lápiz y papel. Fueron diseñadas para cumplir con los requisitos legales estatales y federales relacionados con la implementación del currículo estatal y de su programa de rendición de cuentas. Con ese fin, las pruebas EOG se alinearon estrechamente con el currículo estatal durante su desarrollo. Fueron diseñadas para medir con precisión el conocimiento y las habilidades de estudiantes individuales, así como las de grupos de estudiantes, y para permitir que el estado monitoreara el crecimiento. Las pruebas de lectura EOG fueron diseñadas para aplicarse al final de cada grado en los grados del 3 al 8, durante las últimas tres semanas de clases.

El desarrollo de ítems y la construcción de la prueba se basaron en la IRT empleando el modelo logístico de tres parámetros (3PL). La consistencia interna de las pruebas de comprensión lectora (alfa de Cronbach) se ubicó entre 0,90 y 0,94. Se informó que la fiabilidad test-retest de la prueba de comprensión lectora fue de 0,86, mientras que su validez de criterio y constructo también se refrendó en varios estudios reportados por Sanford (1996).

Las pruebas de lectura NC EOG han sido modificadas periódicamente para reflejar cambios curriculares. Así, se implementó una segunda edición en 2003, seguida de una tercera en 2008. (La edición actual, basada en los Estándares Comunes Básicos Estatales [Common Core State Standards (CCSS)], fue implementada en 2012–13, pero no se refleja en los ejemplos de datos empleados en el presente artículo). Bazemore y Van Dyk (2004) describieron la segunda edición de estas pruebas de lectura, mientras que la tercera edición fue descrita por el Departamento de Instrucción Pública de Carolina del Norte (North Carolina Department of Public Instruction [NCDPI], 2009). Estos

informes técnicos confirman que las ediciones posteriores de las pruebas EOG tienen características de confiabilidad y validez similares a las de la primera edición.

Las pruebas NC EOG han sido objeto de varias auditorías y revisiones técnicas desde su desarrollo inicial, incluyendo las revisiones de evaluación de pares requeridas por la ley NCLB. Siempre se ha reportado que son técnicamente suficientes para los propósitos para los cuales fueron diseñadas. Al menos en una ocasión, fueron seleccionadas como las mejores de todo el país (Princeton Review, 2002).

En este artículo se examinaron los puntajes de lectura EOG de 10 cohortes sucesivas de estudiantes de octavo grado (2000-2009) retrospectivamente durante seis años, a partir de fines del tercer grado hasta el final de octavo. Se construyeron paneles longitudinales con las mediciones de lectura de los estudiantes para los años 1995–2009. Esta compilación de datos emparejados longitudinalmente produjo 10 paneles de cohortes. Los 10 paneles sucesivos representan el crecimiento de la habilidad lectora entre el 3er y el 8vo grado, pero a lo largo de distintos años: 1995–2000, 1996–2001, 1997–2002, 1998–2003, 1999–2004, 2000–2005, 2001–2006, 2002–2007, 2003–2008 y 2004–2009, respectivamente.

Para facilitar la medición del crecimiento, la primera edición de las Pruebas de Comprensión Lectora de Fin de Grado de Carolina del Norte fue diseñada con una escala de desarrollo. Esta iba aproximadamente de 100 a 200 puntos. La escala vertical fue construida mediante un diseño equiparador de ítems comunes.

Cuando en 2003 se puso en operación la segunda edición, esta reflejó las modificaciones realizadas al Curso Estándar de Estudios para la Lectura (Standard Course of Study for Reading) y, por lo tanto, se diseñó una nueva escala de desarrollo de 200 a 300 puntos para diferenciarla de la escala de la primera edición. Se llevó a cabo un estudio de enlace para permitir que se tradujeran los resultados de la segunda edición a la escala de la primera edición con propósitos de rendición de cuentas (por ejemplo, para facilitar el cálculo del avance logrado durante un año de transición hasta tener dos años de datos disponibles en la nueva escala).

Asimismo, cuando en 2008 se puso en operación la tercera edición de las pruebas de lectura, esta reflejó las nuevas modificaciones realizadas al Curso Estándar de Estudios para lectura y lenguaje (Standard Course of Study for reading and language arts) y, por lo tanto, se diseñó una nueva escala de desarrollo lector, esta vez de 300 a 400 puntos para diferenciarla de las escalas de la primera y la segunda edición. Un nuevo estudio de enlace permitió traducir los resultados de la tercera edición a la escala de la segunda edición con el objetivo de rendir cuentas.

En su investigación, Williamson (2014) aprovechó los estudios de enlace realizados entre ediciones para ubicar todos los puntajes de un panel de cohortes dentro de la escala de la primera edición. Luego, dado que el Departamento de Instrucción Pública del Estado (NCDPI) había enlazado la primera edición de las Pruebas de Comprensión Lectora de Final de Grado de Carolina del Norte con el Marco Lexile para la Lectura, todos los puntajes de la primera edición (tanto los puntajes originales de la primera edición como los puntajes convertidos de las escalas de la segunda y la tercera edición) fueron expresados como medidas Lexile para sus análisis de crecimiento. Mirando retrospectivamente, se puede observar que el Estado de Carolina del Norte aprovechó la IRT tradicional para construir sus pruebas y escalas verticales. Luego, en virtud del estudio de enlace Lexile, la escala vertical del estado fue calibrada con respecto a la escala Lexile, la cual utilizó un modelo Rasch (ajustado a la Prueba de Enlace Lexile) y la ecuación de especificación de constructo para calibrar los continuos de habilidad personal y de complejidad de tarea. De este modo, la escala de lectura NC EOG fue ubicada en un marco interpretativo conjunto que posee objetividad general.

**Modelamiento del crecimiento de los estudiantes.** Existen cinco propiedades que considero esenciales para la medición óptima del crecimiento de los estudiantes. En primer lugar, la escala de medición debe ser *unidimensional*. Es decir, la escala debiera medir solo un constructo consistentemente a lo largo del tiempo. Si esto no se cumple, los investigadores no pueden establecer claramente qué es lo que crece desde una ocasión a otra. Así, por ejemplo, los investigadores pueden medir el progreso de la habilidad lectora individual a través del tiempo si tienen una medición de habilidad lectora que mantenga su validez durante el ciclo vital de un individuo.

En segundo lugar, la escala de medición debe ser *continua* en el sentido matemático usual. Prefiero usar una función continua y diferenciable para describir el proceso de crecimiento, porque deseo modelar no solo cómo cambia el estado con el tiempo, sino que también la velocidad y la aceleración de dicho crecimiento. Se debe tener en cuenta que, incluso si el crecimiento no es continuo, es ventajoso que la escala de medición sea continua. Un proceso de crecimiento discontinuo puede modelarse con una escala continua, pero un proceso de crecimiento continuo no puede modelarse adecuadamente con una escala categórica.

En tercer lugar, la escala de medición debe tener en cuenta el *desarrollo*. Es decir, la escala debe incluir los grados escolares o la edad cronológica. Esto permite recolectar mediciones comparables y repetidas de un individuo, a medida que avanza en su educación escolar o a lo largo de su proceso de maduración.

En cuarto lugar, la escala de medición debe ser de *intervalos iguales*. Esto significa que la unidad de la escala debe tener la misma cuantificación (es decir, debe representar la misma cantidad del constructo) en cualquier lugar de la escala. Para las escalas IRT, que tengan intervalos iguales en ocasiones se interpreta como que la escala es invariable ante la transformación lineal (Yen & Fitzpatrick, 2006). La mayoría de los investigadores aceptaría que estas cuatro propiedades bastan para medir el crecimiento. Sin embargo, existe una quinta propiedad que considero necesaria.

La quinta propiedad importante de medición es la *invariancia de la unidad de la escala*. Otra forma de decir esto es que la medición posee objetividad general (Stenner & Stone, 2010). En términos simples, esto quiere decir que el tamaño de la unidad de la escala es absoluto. No cambia solo porque se use una prueba de lectura distinta. No necesita referencias a ningún individuo o individuos específico(s). La objetividad general se obtiene mediante una base teórica o anclando la escala en un contexto del mundo real (por ej., de forma análoga a cómo se ancló el metro en términos del largo del arco de meridiano que pasa por París).

Como se describe en la sección anterior, la escala vertical NC EOG fue enlazada con el Marco Lexile para la Lectura, el cual utiliza un modelo de medición Rasch con una ecuación de especificación de constructo para ubicar tanto a las personas como a los textos dentro de una escala común, que está anclada en dos puntos en un continuo bien definido de complejidad textual (Stenner et al., 2007). En consecuencia, con respecto a los ejemplos que presentaré más adelante, puedo afirmar que la escala posee las cinco propiedades delineadas anteriormente, además de proveer medición conjunta tanto de personas como de textos.

Los siguientes ejemplos de crecimiento estudiantil emplean los resultados de investigaciones previas, descritas en detalle por Williamson (2014). Basada en las recomendaciones entregadas por Singer y Willett (2003), la primera fase de la estrategia de análisis de Williamson involucraba la construcción de cuadros exploratorios de los datos longitudinales individuales, suavizando no paramétricamente los datos de tendencia con splines, ajustando modelos de regresión de mínimos cuadrados ordinarios (por ej., modelos de transformación lineales, cuadráticos y en función del logaritmo del tiempo) a los datos individuales, y observando la distribución de estadísticas de ajuste en la población. Además, Williamson examinó tendencias globales en el desempeño en los grados del 3 al 8 en cada cohorte sucesiva de estudiantes. De esta manera, confirmó que la misma función de desarrollo parecía ser apropiada tanto para los datos globales como para los individuales. Las tendencias globales produjeron una visión preliminar de cómo había cambiado el crecimiento a través del tiempo.

En la segunda fase de su estrategia analítica, los análisis exploratorios de Williamson fueron seguidos por el modelamiento estadístico formal de los datos. Ajustó una secuencia de modelos multinivel para cambio, los cuales fueron ideados para confirmar la forma funcional esperada y generar caracterizaciones matemáticas específicas del crecimiento. Así, se utilizó un modelo paramétrico para caracterizar matemáticamente la naturaleza del desarrollo del crecimiento individual desde el final del tercer grado hasta el final del octavo grado. Una vez que el modelo de crecimiento hipotético fue ajustado a los datos de los paneles de cada cohorte, Williamson (2014) obtuvo: (a) un modelo ajustado para resumir el crecimiento, (b) estimaciones de parámetros asociados con sus errores estándar y (c) estimaciones de la variabilidad del crecimiento individual de los estudiantes en el estado. Las estimaciones de parámetros obtenidas por Williamson son directamente interpretables en términos de las características del

crecimiento (magnitud, velocidad y aceleración). Al examinar la forma paramétrica a través de paneles de cohortes sucesivas, logró explicar patrones observados en las características del crecimiento a través del tiempo en Carolina del Norte. Las estimaciones paramétricas de efectos fijos reportadas por Williamson (2014) fueron los resultados básicos empleados en las ilustraciones de crecimiento estudiantil entregadas en el presente artículo.

### Propósito y metodología

Deseo demostrar, mediante una serie de ejemplos, cómo una comprensión conjunta del crecimiento lector de los estudiantes en relación con estándares de complejidad textual puede ser útil para informar el debate sobre políticas dirigidas a mejorar el crecimiento estudiantil en el futuro. Cuatro puntos centrales orientan este propósito.

En primer lugar, en EE.UU. ha emergido una nueva política que conforma una plataforma para medir el crecimiento de los estudiantes. Desde la aprobación de la ley NCLB, es posible documentar el crecimiento lector histórico de los estudiantes de algunos estados debido a los requisitos legales para las pruebas. La ley NCLB exige a los estados que realicen pruebas desde el 3er al 8vo grado y en un nivel secundario. Aunque la ley no indica que los estados deban emplear escalas de desarrollo capaces de entregar una mejor medición del crecimiento estudiantil, algunos estados sí utilizaron dichas escalas. En estados que mantuvieron datos longitudinales e implementaron mediciones suficientes, fue posible medir crecimiento en lectura y matemáticas desde el final del 3er grado hasta finales del 8vo (Williamson, Thompson, & Baker, 2007). Al menos un estado (Carolina del Norte) también empleó una escala suplementaria de medición conjunta de modo de poder reportar, con una escala común, tanto la habilidad lectora como la complejidad textual. En esta situación, el crecimiento lector histórico de los estudiantes entrega un contexto para entender cuán desafiantes podrían ser en la práctica los estándares emergentes de complejidad textual de los Estándares Comunes Básicos Estatales (CCSS).

En segundo lugar, el Centro para las Buenas Prácticas de la Asociación Nacional de Gobernadores y el Consejo de Jefes de Escuelas Estatales (National Governors Association [NGA], & Council of Chief State School Officers [CCSSO], 2012) propusieron rangos cuantitativos de complejidad textual para bandas específicas de grados para así animar a los educadores a exigir a sus alumnos leer textos más complejos. Sus Estándares Comunes Básicos Estatales (CCSS) recomendaban incrementar la complejidad desde Kindergarten hasta el grado 12 como una estrategia para reducir la brecha entre la capacidad de lectura en la escuela secundaria y la complejidad de los materiales de lectura post-secundarios (Williamson, 2008). Williamson, Fitzgerald y Stenner (2013) discuten algunas de las implicaciones de los estándares de complejidad textual.

En tercer lugar, el conocimiento sobre el crecimiento histórico inspira y motiva trayectorias hipotéticas de crecimiento compatibles con el logro de la preparación necesaria para la universidad y el mundo laboral al término del grado 12. Con un modelo de crecimiento paramétrico adecuado, las trayectorias alternativas pueden ser examinadas analíticamente en términos de tres características básicas del crecimiento: magnitud, velocidad y aceleración.

En cuarto lugar, esta estrategia de considerar trayectorias alternativas sienta las bases de una conversación más rica entre los educadores acerca de cómo las políticas y prácticas educativas pueden hacer que los estudiantes estén mejor preparados para el mundo post-secundario. Williamson, Fitzgerald y Stenner (2014) exploran esta conversación en profundidad.

Los ejemplos destacados en el presente artículo ilustran las conexiones entre estas cuatro ideas focales. Con ese fin, se presentará una serie de figuras y se describirá un útil experimento mental. Comienzo por el crecimiento de los estudiantes.

### Ejemplo de crecimiento de los estudiantes

En la Figura 1 vemos un ejemplo de medición del crecimiento de los estudiantes. El eje horizontal usa el grado dentro de la escuela como escala temporal. Dado que los estudiantes rendían pruebas

anuales en los grados del 3 al 8 durante las últimas tres semanas del año escolar, los numerales del eje horizontal representan momentos separados aproximadamente un año el uno del otro, y corresponden a los grados marcados en el eje. El eje vertical cuantifica la habilidad lectora estimada asociada con las curvas de crecimiento mostradas en el gráfico. Aunque la escala vertical va desde bajo 0L hasta sobre 2000L, solamente muestro la porción de la escala entre 600L y 1200L para así permitir una mejor discriminación al leer el gráfico.

Las 10 curvas de la figura representan el crecimiento promedio estimado para los 10 diferentes grupos de estudiantes que fueron seguidos y medidos a lo largo de seis años de educación escolar (del 3er al 8vo grado). Cada curva representa el logro de los estudiantes durante el mismo grupo de grados (3–8), pero en años diferentes. Por ejemplo, la curva inferior del cuadro corresponde a los primeros años. Los estudiantes asociados con la curva de crecimiento inferior estaban en 3er grado en la primavera de 1995 y avanzaron en sus estudios sin repetir grados hasta llegar al 8vo grado en la primavera de 2000. La curva siguiente representa a los estudiantes que cursaban de 3er a 8vo grado durante los años 1996–2001, y así sucesivamente hasta llegar a los estudiantes asociados con la curva superior, que corresponde a los grados del 3ro al 8vo en los años 2004–2009. Cada grupo de estudiantes fue medido en seis oportunidades. En total, 674.899 estudiantes conformaron estos 10 grupos.

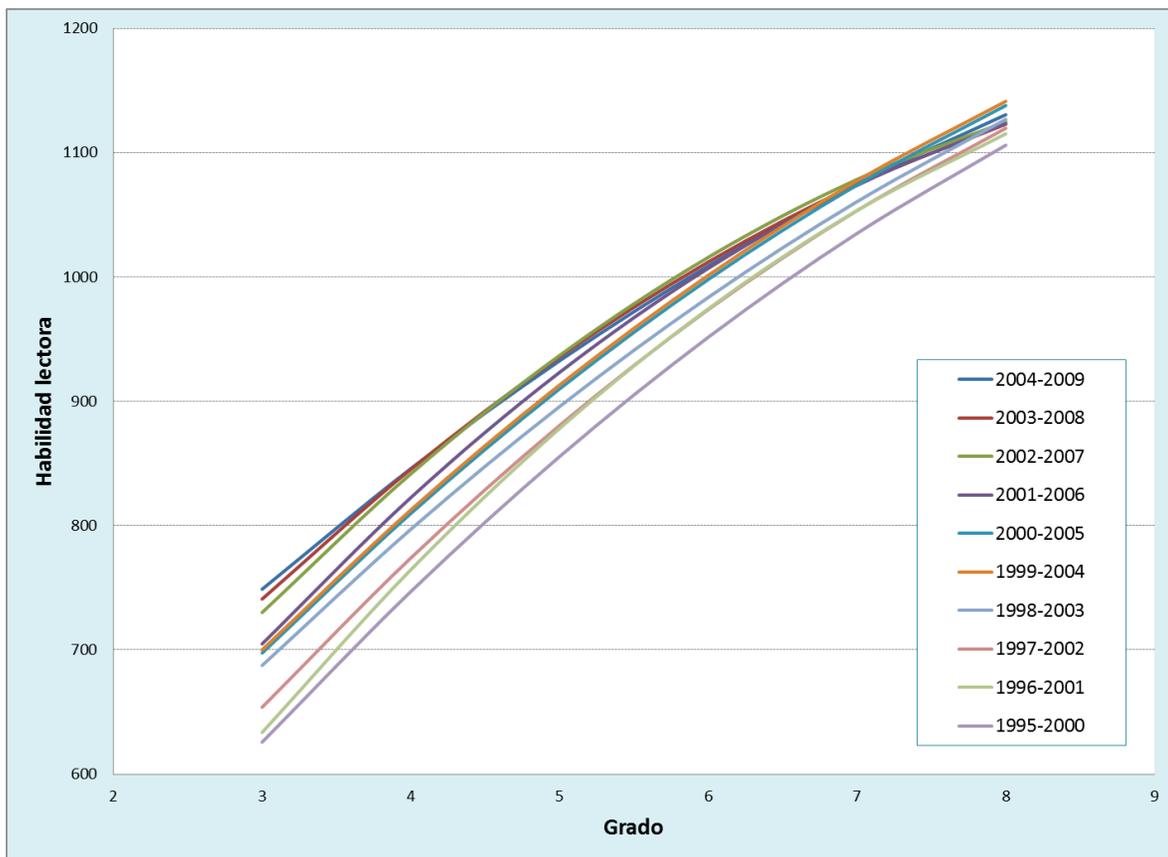


Figura 1. Crecimiento lector promedio en Carolina del Norte para 10 paneles de estudiantes (N = 674.899). Cada curva se basa en seis grupos de datos correspondientes a los grados del 3ro al 8vo durante diferentes años, como indica la leyenda.

Las curvas de crecimiento de la Figura 1 representan el crecimiento promedio de los estudiantes en Carolina del Norte, un estado del sur de EE.UU., durante los años y períodos especificados. Puesto que las curvas se ubican verticalmente una sobre otra en orden cronológico en el eje vertical, es evidente, de acuerdo a la Figura 1, que Carolina del Norte experimentó una mejora sistémica en su logro lector promedio en múltiples grupos de estudiantes educados a lo largo de 1,5 décadas (1995–2009).

También se observa que la velocidad y la aceleración fueron bastante consistentes entre una curva y otra, exceptuando una desaceleración ligeramente mayor en algunas de las curvas más recientes, lo cual se asoció con un desempeño promedio algo menor al final del 8vo grado. A pesar de que la evidencia muestra una desaceleración ligeramente mayor en algunos grupos, el crecimiento promedio parece ser relativamente fuerte y consistente en estos estudiantes.

### Ejemplos de complejidad textual

En la Figura 2 se emplean cuadros de caja con bigotes (box-and-whisker plots) para ilustrar distribuciones de complejidad dentro de los grados. Una vez más, el eje horizontal se calibra en términos de grado, cubriendo los grados del 1 al 12. Un punto adicional, denominado CCR (College and Career Readiness), fue añadido después del grado 12 en el eje horizontal para denotar el mundo post-secundario de la *preparación para los estudios superiores y el ámbito laboral*. El eje vertical es una escala de complejidad textual que va de 0L a 1600L. Valores mayores en esta escala indican una dificultad o una complejidad textual más alta.

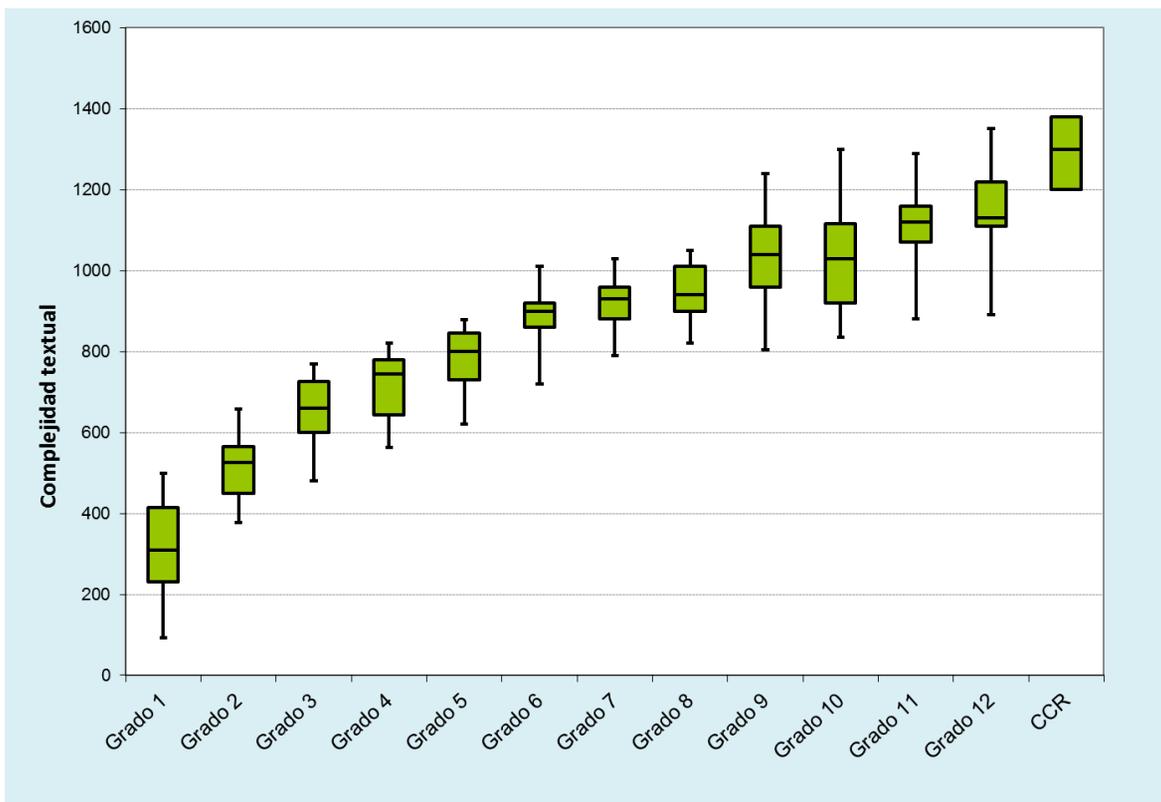


Figura 2. Distribuciones empíricas de complejidad textual por grado y con respecto a la preparación para los estudios superiores y el mundo laboral (CCR). Los cuadros de caja con bigotes representan los percentiles 5, 25, 50, 75 y 95 de las distribuciones de textos.

En el cuadro, las cajas individuales representan el rango intercuartílico (50% intermedio) de complejidad textual de los libros de texto usados en ese grado específico o en el mundo post-secundario. La línea dentro de la caja representa la complejidad textual media del grado. Los bigotes bajan hasta el percentil 5 de los textos y suben hasta el percentil 95. Por lo tanto, la distancia entre los extremos de los bigotes representa la complejidad textual del 90% de los textos asociados con ese grado. La caja y los bigotes, en conjunto, representan la variabilidad dentro de los grados en cuanto a la complejidad textual de los libros de texto usados comúnmente en las escuelas públicas estadounidenses.

El cuadro revela una tendencia ascendente en la complejidad textual media desde los grados menores a los mayores, incluyendo también el material de lectura post-secundario. La complejidad textual media aumenta más rápidamente durante los primeros grados (por ej., del 1 al 6) y disminuye en los años posteriores. Aunque existe una variabilidad considerable dentro de los grados en cuanto a complejidad textual y distintos niveles de superposición en las distribuciones de complejidad textual de grados adyacentes, el gráfico indica un gran aumento en la magnitud de la complejidad textual para casi todos los textos durante el desarrollo de los estudiantes a medida que se acercan a la adultez. Sin embargo, el aumento de la complejidad textual media entre el final del grado 12 y la etapa post-secundaria es de especial interés, ya que cuantifica la brecha en habilidad lectora necesaria para leer material post-secundario en comparación con textos de educación secundaria.

Dado que la elección de libros de texto para la enseñanza escolar es una política educativa y una decisión pedagógica modificable, es posible imaginar un continuo textual en el cual los libros de texto de los grados 1 al 12 estén más alineados con la complejidad textual de los materiales de lectura post-secundarios. Al desplazar las distribuciones de complejidad textual para los grados del 1 al 12 hacia arriba, hasta que la complejidad textual media del grado 12 se alinee con la de los textos post-secundarios, llegamos a un continuo textual *aspiracional*, mostrado en la Figura 3. De hecho, este es el enfoque que subyace a los estándares de complejidad textual de los Estándares Comunes Básicos Estatales (CCSS). Esta relación se evidencia en la Figura 3 al observar los límites superiores e inferiores de los rangos de complejidad de los CCSS (las dos líneas grises que van del grado 3 al grado 12). Sanford-Moore y Williamson (2012) resumieron una estrategia para desplazar un continuo empírico de complejidad textual y así crear un continuo de complejidad textual aspiracional. Los rangos de complejidad textual incluidos en los CCSS fueron descritos por la NGA y el CCSSO (2012).

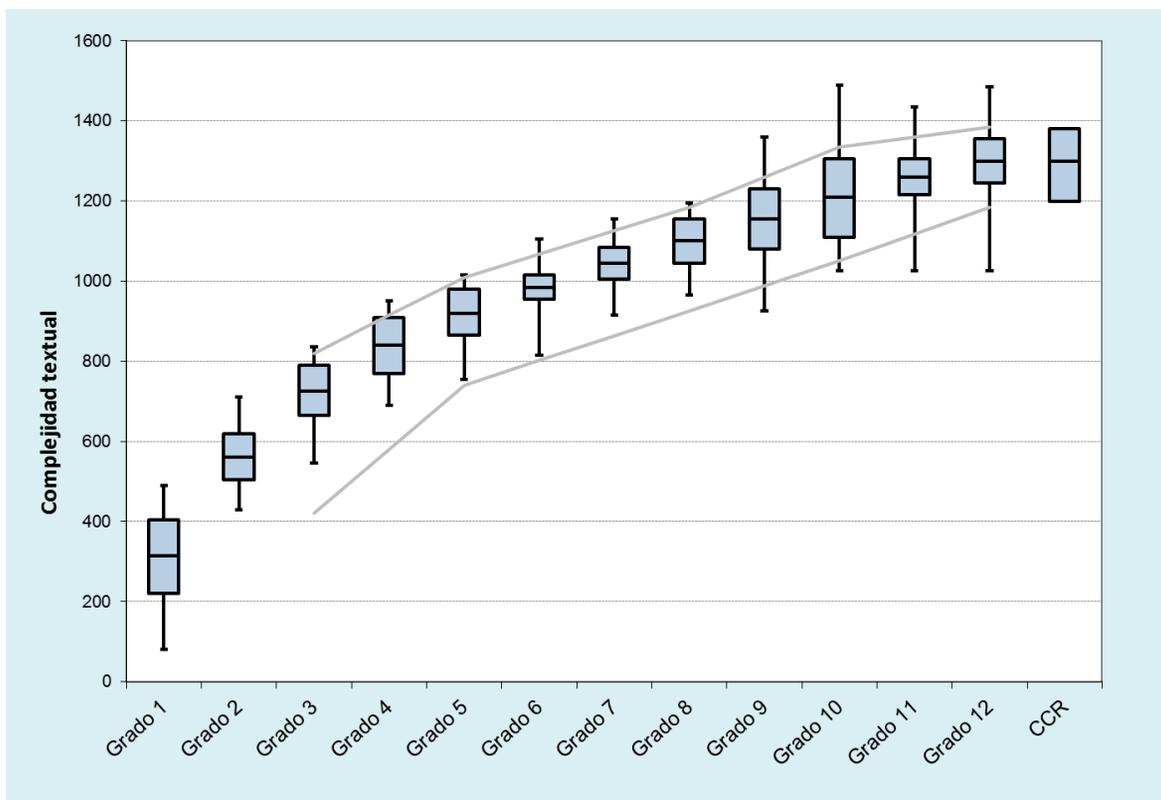


Figura 3. Distribuciones objetivo de textos, rangos de complejidad textual de los Estándares Comunes Básicos Estatales (CCSS) (en gris) y rango intercuartílico de la complejidad textual de textos post-secundarios (CCR). Los cuadros de caja con bigotes representan los percentiles 5, 25, 50, 75 y 95 de las distribuciones de textos.

### La perspectiva conjunta

El poder de combinar mediciones Rasch con un modelo teórico de la complejidad de las tareas se evidencia en la Figura 4. Una vez más, el eje horizontal indica el tiempo (es decir, el final de los grados consecutivos). El eje vertical es exactamente la misma escala empleada en las Figuras de la 1 a la 3. La característica más llamativa de la Figura 4 es que las mediciones cuantitativas tanto de personas como de textos se manifiestan en un gráfico usando una sola escala.

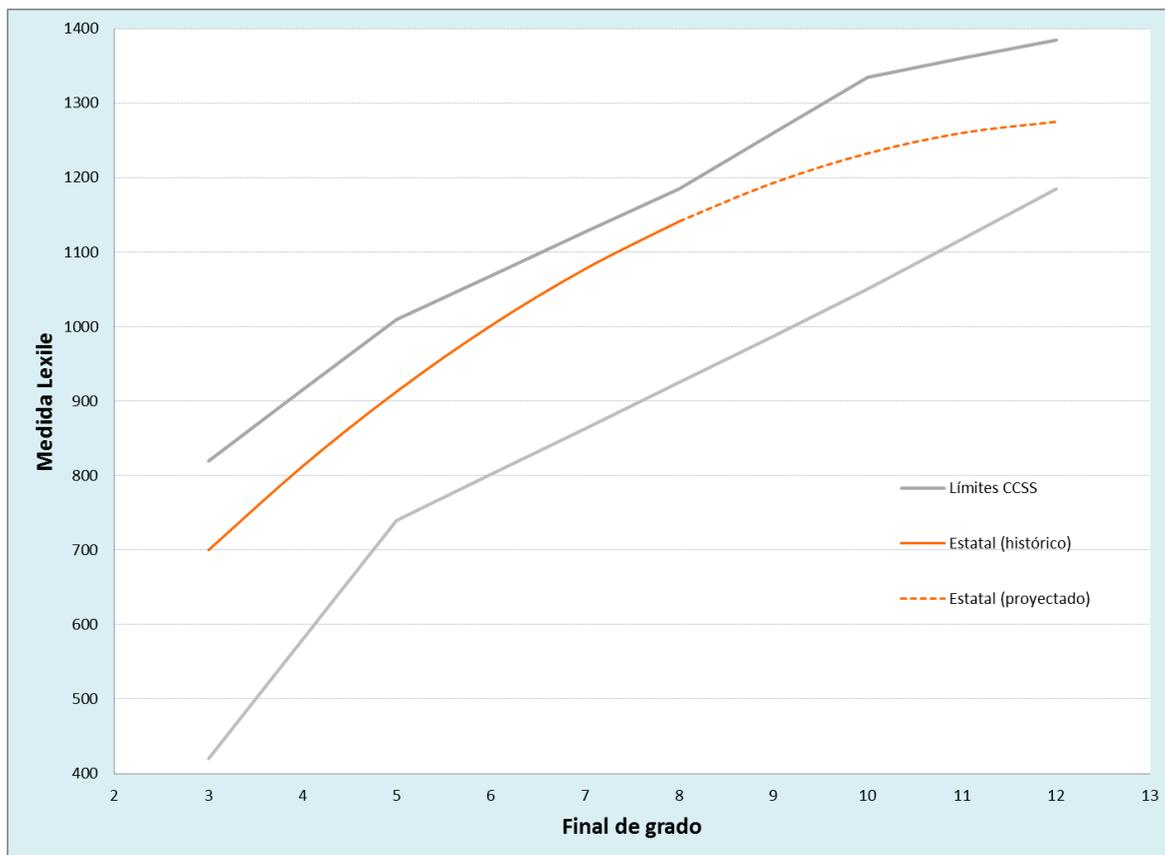


Figura 4. Crecimiento estudiantil promedio y crecimiento estudiantil proyectado de acuerdo a los rangos de complejidad textual de los Estándares Comunes Básicos Estatales (CCSS). La habilidad lectora de los estudiantes y la complejidad textual son cuantificadas con una escala común (Lexile).

De esta manera, dos perspectivas aparentemente dispares (persona y tarea) pueden contrastarse dentro de un marco de referencia común. Con respecto a la Figura 4, los lectores podrán notar que los límites de complejidad textual de los Estándares Comunes Básicos Estatales (CCSS) se muestran desde el grado 3 al 12. Estos son los mismos límites de complejidad textual anteriormente ilustrados en la Figura 3. El segundo aspecto notable de la Figura 4 es la curva global de crecimiento estudiantil. Parte de la curva se muestra mediante una línea continua, mientras que la otra parte aparece como un patrón discontinuo. La porción continua de la curva representa la curva de crecimiento ajustada empíricamente; de hecho, esta curva es una de las 10 que se muestran en la Figura 1 (la de los años 1999–2004). La porción discontinua de la curva es una extrapolación matemática basada en el modelo de crecimiento ajustado.

En la Figura 4 es ahora muy fácil ver la relación entre los estándares de complejidad textual y el crecimiento promedio de los estudiantes, tanto el crecimiento histórico real entre los grados 3 y 8 como el crecimiento proyectado entre los grados 9 y 12, con la condición de que los estudiantes sigan el mismo camino histórico hacia el futuro sin ningún cambio en su curva de crecimiento. La imagen

permite apreciar claramente dos resultados: (a) durante los grados del 3 al 8, los estudiantes que están en el promedio parecen desempeñarse bien en relación con los estándares de complejidad textual de los CCSS, subiendo hasta acercarse al límite superior del rango de complejidad textual hacia fines del 8vo grado; y (b) la desaceleración de la trayectoria histórica implica que los estudiantes podrían incluso quedar debajo de las exigencias de complejidad textual a fines del grado 12, si no ocurre nada que atenúe su desaceleración. Por supuesto, la curva de crecimiento de la Figura 4 representa crecimiento promedio. Las curvas de crecimiento individual probablemente variarían en torno a la curva de crecimiento promedio: algunos estudiantes rendirían mejor en relación con los estándares de complejidad textual pero otros probablemente tendrían resultados muy inferiores al promedio. Esto plantea una pregunta potencialmente relevante. ¿Cómo se puede alterar el crecimiento de los estudiantes para que estos puedan alcanzar mayores niveles de rendimiento en el futuro? Es decir, ¿cómo deberían cambiar las características del crecimiento (magnitud, velocidad, aceleración) para que los estudiantes logren un resultado aspiracional distinto?

### **Implicaciones para los educadores**

Para que los educadores puedan considerar las implicaciones de largo plazo del crecimiento, deben pensar más allá del estado inmediato de un individuo en un punto específico del tiempo. De hecho, deben superar las ideas de corto plazo sobre el crecimiento, como por ejemplo el progreso logrado durante un año escolar dado. En realidad, los educadores deberían ver la trayectoria completa de crecimiento como una entidad dinámica con características específicas que determinan el crecimiento a lo largo del ciclo vital. Más concretamente, es útil enfocarse en características específicas del crecimiento (por ej., magnitud, velocidad y aceleración), que son inherentes al proceso de crecimiento. Con esta perspectiva, es posible refinar preguntas y estrategias pedagógicas para centrarse en características específicas del crecimiento. Por ejemplo, ¿qué característica del crecimiento debe cambiar (y cuánto) para lograr un resultado terminal distinto? La Figura 5 muestra un ejemplo de un experimento mental de este tipo.

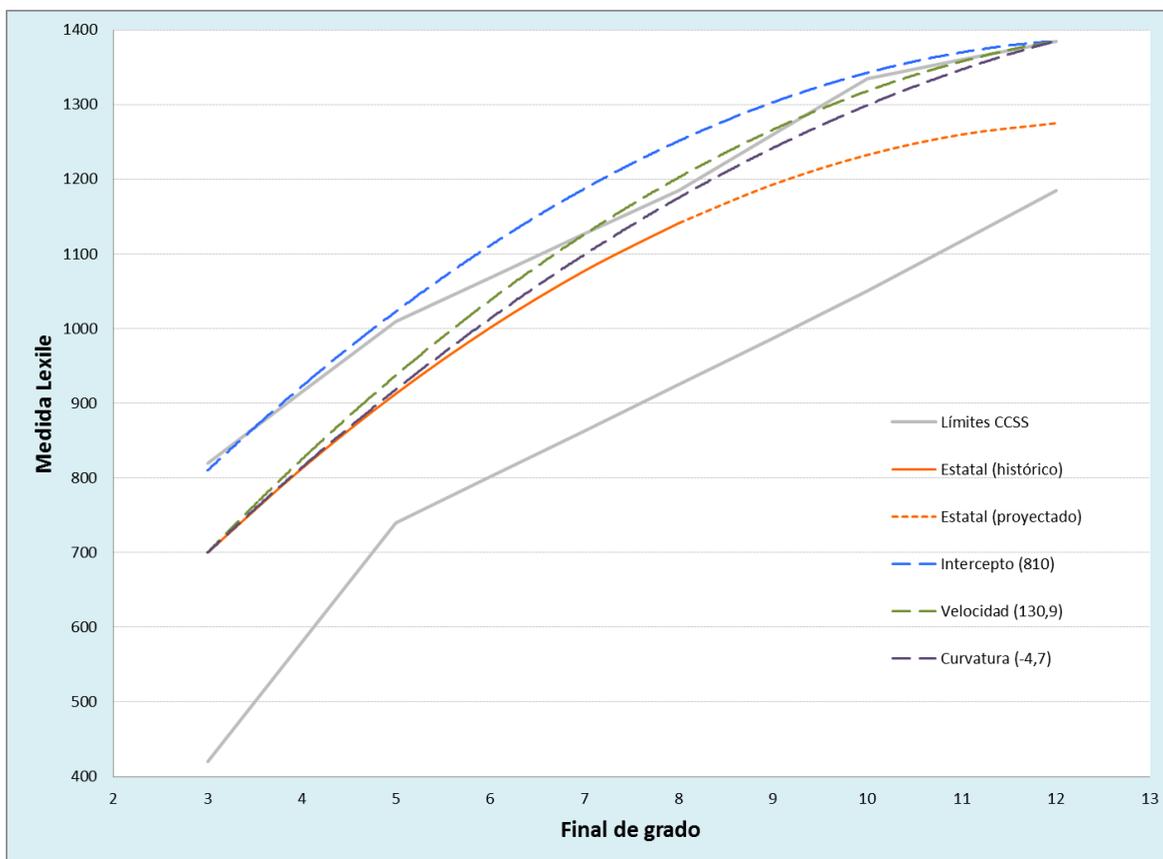


Figura 5. Crecimiento histórico promedio relativo a los rangos de complejidad textual de los Estándares Comunes Básicos Estatales (CCSS), con tres caminos alternativos para alcanzar 1385L. La habilidad lectora de los estudiantes y la complejidad textual son cuantificadas con una escala común (Lexile). Adaptado de «Student Reading Growth Illuminates the Common Core Text-Complexity Standard: Raising Both Bars» de G. L. Williamson, J. Fitzgerald y A. J. Stenner, 2014, *Elementary School Journal*, 115(2), p. 245. © 2014 University of Chicago.

Se debe considerar que la Figura 5 tiene los mismos ejes que la Figura 4 y que los dos elementos que aparecen en esta última (los límites de complejidad textual de los CCSS y la curva de crecimiento promedio) se replican en la Figura 5. Sin embargo, la Figura 5 incluye tres curvas adicionales. Estas curvas entregan ejemplos de tres caminos arquetípicos conducentes a un resultado aspiracional (1385L), que es la meta CCR establecida por los estándares de complejidad textual de los CCSS.

La curva superior de la Figura 5 representa un camino a 1385L que se logra aumentando la elevación de la curva de crecimiento histórico completa en todos los grados. Específicamente, la curva superior tiene un punto inicial más alto (810L) que la curva histórica (700L), pero ambas son paralelas en todos los demás aspectos. La consecuencia de este cambio único es que la nueva curva alcanza el objetivo CCR, mientras que la predicción basada en la curva histórica no lo logra. La curva inmediatamente inferior fue creada cambiando la velocidad inicial de la curva histórica, de modo tal que la nueva curva alcanzara a 1385L. Finalmente, la tercera curva desde arriba fue construida cambiando la tasa de desaceleración de la curva histórica solo lo necesario para alcanzar 1385L al final del grado 12.

Es importante comprender que este experimento mental es una sobresimplificación. Frecuentemente existe una estructura de correlación entre los parámetros de las curvas empíricas de crecimiento, es decir, los parámetros de estado, velocidad y aceleración no son completamente independientes entre sí. Por lo tanto, los investigadores no pueden cambiar solo una de estas tres características sin esperar un cambio concomitante en las demás. Sin embargo, estoy ignorando esa realidad con un fin heurístico. Las tres nuevas trayectorias no son primordiales en sí mismas. Lo que sí importa son las preguntas que plantean

estas tres nuevas trayectorias sobre la práctica educativa y, para esa discusión, es útil pensar sobre cada posibilidad una a una.

Al observar la más alta de estas nuevas curvas, la pregunta pertinente para las políticas y las prácticas educativas es cómo impactar el estado inicial de un grupo de estudiantes. Esta es una pregunta común con la cual los educadores tienen mucha experiencia. La respuesta típica consiste en explorar las opciones para una intervención temprana. La curva inmediatamente inferior implica preguntarse cómo cambiar la velocidad del aprendizaje de los estudiantes. Las respuestas probablemente se centren en el ritmo de la enseñanza durante el lapso de tiempo considerado. La tercera curva nos conduce a buscar una forma de reducir la desaceleración. Esta pregunta rara vez se tiene en cuenta, excepto tal vez en el contexto de programas diseñados para combatir el olvido veraniego en el aprendizaje de los estudiantes.

Empleando los escenarios ilustrados en la Figura 5, Williamson et al. (2014) entregan una discusión profunda de las implicaciones pedagógicas y políticas para intentar modificar la magnitud, velocidad o aceleración del crecimiento de los estudiantes; además, también discuten alternativas e implicaciones posibles en el contexto de la investigación existente sobre la lectura. Al considerar cada escenario alternativo de crecimiento estudiantil, los autores abordaron los siguientes tres puntos basándose en evidencias:

- (a) ¿Qué implicaciones tienen los niveles de desafío de los textos que los estudiantes necesitarían leer en los distintos grados, y cómo se comparan estos niveles de desafío con los límites para cada grado incluidos en los Estándares Comunes Básicos Estatales (CCSS)? (b) En comparación con la curva histórica de crecimiento lector de los estudiantes, ¿qué estudiantes resultarían más afectados por el cambio, y cómo podría la investigación en lectura informar nuestra consideración de dicho cambio? (c) ¿Cuáles son las implicaciones para el currículo, la enseñanza o las políticas educativas? (p. 246).

El análisis de Williamson et al. (2014) es un ejemplo de cómo los educadores pueden abordar intelectualmente las implicaciones sustanciales del crecimiento lector de los estudiantes en el contexto de los estándares de complejidad textual.

### Discusión

Existen varios beneficios adicionales derivados de la perspectiva metodológica descrita en el presente artículo. Primero, esta perspectiva motiva una conversación sobre crecimiento académico que está fundamentalmente anclada en un proceso de medición sólido. Segundo, esta perspectiva insta a los educadores a adoptar un enfoque con respecto al crecimiento de los estudiantes de largo plazo y que considere todo el ciclo vital, basado en recolectar y analizar datos longitudinales. Tercero, este enfoque fomenta que los educadores conceptualicen el crecimiento como un proceso de desarrollo dinámico. Estos beneficios permiten conocer mejor el crecimiento de los estudiantes, lo que a su vez puede informar las políticas y prácticas del ámbito educacional.

Las lecciones más importantes contenidas en el presente artículo tal vez sean las referidas a los profesionales de la medición educacional. Para medir el crecimiento académico de forma óptima se necesita una escala unidimensional, continua, de intervalos iguales, de desarrollo e invariante con respecto a la ubicación y al tamaño de la unidad. Para el constructo de habilidad de lectura, estas características pueden lograrse combinando la medición Rasch con una especificación operativa del constructo de lectura (en términos del continuo de tareas) y anclando la escala resultante en dos puntos del continuo de complejidad textual. De esta manera, las personas y los textos se llevan a una escala común y de desarrollo, la cual posee objetividad general.

Vale la pena recalcar que este enfoque no necesariamente debe restringirse al constructo de lectura. Por ejemplo, si la dificultad de las habilidades y conceptos matemáticos (o de otro tipo) pudiesen cuantificarse, y si se demostrase que están fuertemente relacionados con las dificultades de un tipo específico de ítem de referencia, entonces la habilidad matemática de los estudiantes y la dificultad de las tareas matemáticas podrían escalarse conjuntamente usando el modelo Rasch. Esto produciría una medición conjunta y sustentada teóricamente para las matemáticas (u otros constructos), similar a la demostrada aquí para la habilidad de lectura.

El presente artículo se centró en cómo el crecimiento de los estudiantes entrega un contexto para considerar las implicaciones educacionales de la complejidad textual. De hecho, con el enfoque aquí presentado, la complejidad textual también podría entregar un contexto para comprender el crecimiento de los estudiantes. Este es el doble beneficio de la medición conjunta. Cada faceta (personas, tareas) puede aportar un contexto que ayude a entender la otra.

El artículo original fue recibido el 10 de noviembre de 2014

El artículo revisado fue recibido el 17 de abril de 2015

El artículo fue aceptado el 13 de mayo de 2015

## Referencias

- Alvermann, D. E., Hinchman, K. A., Moore, D. W., Phelps, S. F., & Waff, D. R. (Eds.). (2006). *Reconceptualizing the literacies in adolescents' lives*. (2<sup>nd</sup> ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Alvermann, D. E., Unrau, N. J., & Ruddell, R. B. (Eds.). (2013). *Theoretical models and processes of reading*. (6<sup>th</sup> ed.). Newark, DE: International Reading Association.
- Bazemore, M., & Van Dyk, P. B. (2004). *North Carolina reading comprehension tests: Technical report*. (2<sup>nd</sup> ed.). (Borrador citable). Raleigh, NC: North Carolina Department of Public Instruction.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bormuth, J. R. (1969). *Development of readability analyses*. (Final Report, Project No. 7-0052, Contract No. OEG-3-7-070052-0326). Washington DC: US Office of Education, Bureau of Research, US Department of Health, Education and Welfare.
- Brennan, R. L. (Ed.). (2006). *Educational measurement*. (4<sup>th</sup> ed.). Westport, CT: American Council on Education and Praeger Publishers.
- Burdick, H., & Stenner, A. J. (1996). Theoretical prediction of test items. *Rasch Measurement Transactions*, 10(1), 475.
- Gulliksen, H. (1950). *Theory of mental tests*. Nueva York: Wiley. Reimpreso por Lawrence Erlbaum Associates, Hillsdale, NJ, 1987.
- Holland, P. W., & Dorans, N. J. (2006). Linking and equating. En R. L. Brennan (Ed.), *Educational measurement*. (4<sup>th</sup> ed., pp. 187-220). Westport, CT: American Council on Education and Praeger Publishers.
- Kendeou, P., van den Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research & Practice*, 29(1), 10-16. doi: 10.1111/ldrp.12025
- Koslin, B. L., Zeno, S., & Koslin, S. (1987). *The DRP: An effectiveness measure in reading*. Nueva York: The College Entrance Examination Board.
- Legendre, A. M. (1805). *Nouvelles méthodes pour la détermination des orbites des comètes*. Paris: Courcier.
- Lord, F. M. (1952). A theory of test scores. *Psychometric Monograph*, 7. Richmond, VA: Psychometric Corporation.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- National Governors Association (NGA), & Council of Chief State School Officers (CCSSO) (2012). *Supplemental information for Appendix A of the Common Core State Standards for English language arts and literacy: New research on text complexity*. Washington, DC: Autor. Recuperado de <http://www.corestandards.org/resources>
- Nelson, J., Perfetti, C., Liben, D., & Liben, M. (2012). *Measures of text difficulty: Testing their predictive value for grade levels and student performance*. Washington, DC: Council of Chief State School Officers (CCSSO). Recuperado de [http://www.ccsso.org/Documents/2012/Measures%20ofText%20Difficulty\\_final.2012.pdf](http://www.ccsso.org/Documents/2012/Measures%20ofText%20Difficulty_final.2012.pdf)
- No Child Left Behind (NCLB) Act of 2001. *Public Law N° 107-110, § 1, 115 Stat. 1425* (2002).
- North Carolina Department of Public Instruction (NCDPI) (2009). *North Carolina Reading Comprehension Tests: Technical report*. (3<sup>rd</sup> ed.). Raleigh, NC: Autor. Recuperado de <http://www.ncpublicschools.org/docs/accountability/testing/reports/eogreadingtechman3.pdf>
- Pearson, P. D. (2004). The reading wars. *Educational Policy*, 18(1), 216-252. Recuperado de [http://www.corwin.com/upm-data/31886\\_Article1.pdf](http://www.corwin.com/upm-data/31886_Article1.pdf)
- Princeton Review (2002). *Testing the testers 2002: An annual ranking of state accountability systems*. Nueva York: Autor.
- Questar Assessment, Inc. (2012). *Degrees of Reading Power (DRP) Program*. Brewster, NY: Autor. Recuperado de <http://www.questarai.com/Products/DRPProgram/Pages/default.aspx>
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Dinamarca: Nielson and Lydiche (for Danmarks Paedagogiske Institut).
- Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 19, 49-57. doi: 10.1111/j.2044-8317.1966.tb00354.x

- Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. En M. Blegvad (Ed.), *The Danish yearbook of philosophy* (pp. 58-94). Copenhagen: Munksgaard. Recuperado de <http://www.rasch.org/memo18.htm>
- Reckase, M. D. (2009). Logical units for achievement test scores. *NCME Newsletter*, 17(1), 1-15. Recuperado de [http://ncme.org/default/assets/File/pdf/newsletter/vol\\_17\\_num\\_1.pdf](http://ncme.org/default/assets/File/pdf/newsletter/vol_17_num_1.pdf)
- Sanford, E. E. (1996). *North Carolina end-of-grade tests technical report # 1: Reading comprehension, mathematics*. Raleigh, NC: Department of Public Instruction.
- Sanford-Moore, E. E., & Williamson, G. L. (2012). *Bending the text-complexity curve to close the gap* (MetaMetrics Research Brief). Durham, NC: MetaMetrics, Inc.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Nueva York: Oxford University Press.
- Stenner, A. J., Burdick, H., Sanford, E. E., & Burdick, D. S. (2007). *The Lexile framework for reading technical report*. Durham, NC: MetaMetrics, Inc.
- Stenner, A. J., Sanford-Moore, E., & Williamson, G. L. (2012). *The Lexile® framework for reading quantifies the reading ability needed for «College & Career Readiness»* (MetaMetrics Research Brief). Durham, NC: MetaMetrics.
- Stenner, A. J., Smith, M., & Burdick, D. S. (1983). Toward a theory of construct definition. *Journal of Education Measurement*, 20, 305-316. doi: 10.1111/j.1745-3984.1983.tb00209.x
- Stenner, A. J., & Stone, M. (2010). Generally objective measurement of human temperature and reading ability: Some corollaries. *Journal of Applied Measurement*, 11(3), 244-252.
- Thissen, D., & Wainer, H. (Eds.). (2001). *Test scoring*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Williamson, G. L. (2008). A text readability continuum for postsecondary readiness. *Journal of Advanced Academics*, 19, 602-632. doi: 10.4219/jaa-2008-832
- Williamson, G. L. (2014). *Measuring and modeling individual academic growth: Methodological foundations for educational applications in the 21<sup>st</sup> century*. (Manuscrito no publicado).
- Williamson, G. L., Fitzgerald, J., & Stenner, A. J. (2013). The Common Core State Standards' quantitative text-complexity trajectory: Figuring out how much complexity is enough. *Educational Researcher*, 42(2), 59-69. doi: 10.3102/0013189X12466695
- Williamson, G. L., Fitzgerald, J., & Stenner, A. J. (2014). Student reading growth illuminates the Common Core text-complexity standard: Raising both bars. *Elementary School Journal*, 115(2), 230-254. doi: 10.1086/678295
- Williamson, G. L., Koons, H., Sandvik, T., & Sanford-Moore, E. (2012). *The text complexity continuum in grades 1-12* (MetaMetrics Research Brief). Durham, NC: MetaMetrics.
- Williamson, G. L., Thompson, C. L., & Baker, R. F. (2007). *North Carolina's growth in reading and mathematics*. Trabajo presentado en la 2007 Annual Meeting of the American Educational Research Association (AERA), Chicago, IL., EE.UU.
- Yen, W. M., & Fitzpatrick, A. R. (2006). Item response theory. En R. L. Brennan (Ed.), *Educational measurement* (4<sup>th</sup> ed., pp. 64-110). Westport, CT: American Council on Education and Praeger Publishers.