

Estudio de Alineamiento del Currículo de Lengua y Literatura, y Matemática con las Pruebas Ser Bachiller, Ecuador

MIGUEL ÁNGEL HERRERA PAVO^I

JANNETH FERNANDA FIGUEROA CHÁVEZ^{II}

HILDA CAROLINA DÍAZ RUBIANO^{III}

JOSÉ DANIEL ESPINOSA RODRÍGUEZ^{IV}

<http://dx.doi.org/10.22347/2175-2753v11i33.2366>

Resumen

Empleando el método Webb, el presente estudio realiza un análisis del alineamiento de a) las evaluaciones estandarizadas Ser Bachiller, régimen Costa 2017 y 2018, con b) el currículo de los niveles de educación obligatoria del Ecuador (2016), en las áreas de Lengua y Literatura, y Matemática. El trabajo de codificación y análisis tanto de los contenidos del currículo como de las pruebas de evaluación fue realizado por un panel de ocho expertos de cada área. Entre las conclusiones, cabe destacar un alineamiento muy débil en el caso de Lengua y Literatura, donde se excluyen de manera reiterada contenidos relevantes y el nivel de demanda cognitiva de la prueba es menor al planteado en el currículo; y un alineamiento más adecuado en el caso de Matemática, si bien la prueba no logra cubrir de manera oportuna los contenidos del currículo y falla en representar apropiadamente el bloque curricular de Geometría y medida. Las dos ediciones de la prueba analizadas, 2017 y 2018, revelan que de un año a otro no hay variación significativa en los contenidos seleccionados para la evaluación.

Palabras clave: Alineamiento. Pruebas estandarizadas. Estándares de aprendizaje. Currículo. Evaluación del aprendizaje.

Submetido em: 04/06/2019

Aprovado em: 07/10/2019

^I Universidad Andina Simón Bolívar (UASB-E), Quito, Ecuador; <http://orcid.org/0000-0002-0321-7235>; e-mail: miguel.herrera.p@uasb.edu.ec

^{II} Universidad Andina Simón Bolívar (UASB-E), Quito, Ecuador; <http://orcid.org/0000-0003-1911-8175>; e-mail: fercitafigueroa@hotmail.com

^{III} Universidad Andina Simón Bolívar (UASB-E), Quito, Ecuador; <http://orcid.org/0000-0003-3977-4188>; e-mail: hcarolinadiaz@gmail.com

^{IV} Universidad Andina Simón Bolívar (UASB-E), Quito, Ecuador; <http://orcid.org/0000-0002-6457-4922>; e-mail: danielespinosa07@gmail.com

Study of the Alignment of the Curriculum of Language and Literature, and Mathematics with Ser Bachiller Tests, Ecuador

Abstract

Using the Webb method this study analyzes the alignment of a) the standardized assessments of Ser Bachiller Costa 2017 and 2018 regime, with b) the curriculum of compulsory education levels of Ecuador (2016), in the areas of Language and Literature, and Mathematics. The work of coding and analyzing the content of the curriculum and the assessment tests was carried out by a panel of eight experts from each area. Among the conclusions, it is worth mentioning a very weak alignment in the case of Language and Literature, where relevant content is repeatedly excluded and the level of cognitive demand of the test is lower than proposed in the curriculum; and a more appropriate alignment in the case of Mathematics, although the test does not timely cover curriculum content and does not adequately represent the geometry and measurement curriculum. The two editions of the analyzed test, 2017 and 2018, reveal that from one year to the next there is no significant variation in the contents selected for the evaluation.

Keywords: Alignment. Standardized Tests. Learning Standards. Curriculum. Learning assessment.

Estudo do Alinhamento do Currículo de Língua e Literatura, e Matemática com os Testes Ser Bachiller, Equador

Resumo

Utilizando o método de Webb, o presente estudo faz uma análise do alinhamento de a) as avaliações padronizadas do regime de Ser Bachiller, Costa 2017 e 2018, com b) o currículo dos níveis de escolaridade obrigatória do Equador (2016), nas áreas de Língua e Literatura, e Matemática. O trabalho de codificação e análise do conteúdo do currículo e dos testes de avaliação foi realizado por um painel de oito especialistas de cada área. Entre as conclusões, vale destacar um alinhamento muito fraco no caso da Língua e da Literatura, onde conteúdos relevantes são repetidamente excluídos e o nível de demanda cognitiva do teste é inferior ao proposto no currículo; e um alinhamento mais apropriado no caso da Matemática, embora o teste não cubra de forma oportuna o conteúdo do currículo e não represente adequadamente o bloco curricular de Geometria e medição. As duas edições do teste analisado, 2017 e 2018, revelam que de um ano para o outro não há variação significativa nos conteúdos selecionados para a avaliação.

Palavras-chave: Alinhamento. Testes Padronizados. Padrões de Aprendizagem. Currículo. Avaliação da aprendizagem.

Introducción

En los últimos años, han proliferado las evaluaciones estandarizadas de aprendizajes y se han convertido en un elemento clave de la toma de decisiones en política pública, así como en un importante instrumento de opinión sobre la calidad del sistema educativo. Ecuador ha participado en las pruebas promovidas desde el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) y, recientemente, también se ha involucrado en el Programa Internacional para la Evaluación de Estudiantes (PISA) para el Desarrollo; además, el Instituto Nacional de Evaluación (INEVAL) realiza las pruebas Ser. No obstante, sin restarles relevancia, es preciso tener en cuenta que estos tipos de evaluaciones no están alineadas con el currículo escolar, en el caso de las internacionales, o no se ha verificado su alineamiento, en el caso de las nacionales. En este sentido, creemos oportuno determinar cómo el currículo oficial se alinea con las pruebas externas nacionales, constatando los acuerdos entre lo que se evalúa y lo prescrito en el currículo; con esta finalidad, este estudio plantea el análisis del alineamiento en las áreas de Lengua y Literatura, y de Matemática, para el nivel de bachillerato.

Un marco de referencia para los estudios de alineamiento

La evaluación estandarizada resulta atractiva para los responsables de las políticas educativas públicas porque es relativamente económica en comparación con la introducción de posibles mejoras en los programas educativos, puede realizarse de manera externa, se implementa rápidamente y ofrecen resultados palpables (LINN, 2000). Este atractivo está claramente relacionado con su repercusión en los medios de comunicación, y presta una relevancia a la evaluación externa que, más que propiciar decisiones políticas informadas, favorece una suerte de legitimación normativa basada en sus resultados. Este mecanismo tiene especial importancia para quienes pretenden incidir en una determinada visión homogeneizadora de la educación que desestima el hecho de que las prácticas educativas deben estar mediadas social y culturalmente (NORMAND, 2003).

Podemos decir que el objetivo del acto educativo que tiene lugar en el aula y que resume el currículo en acción es que los estudiantes construyan.

Sistemas complejos, personales, de comprensión y de actuación;
combinaciones personales de conocimientos, habilidades,

emociones, actitudes y valores que orientan la interpretación, la toma de decisiones y la actuación de los individuos. (PÉREZ GÓMEZ, 2012, p. 145).

No obstante, las pruebas estandarizadas limitan la evaluación a aquellos componentes de estos sistemas complejos que pueden expresarse a través del instrumento de evaluación que desarrollan: conocimientos y capacidades, dejando al margen aspectos emotivos, actitudinales y valorativos del desarrollo humano (PÉREZ GÓMEZ; SOTO GÓMEZ, 2011, p. 175).

Por otro lado, las evaluaciones estandarizadas que se desarrollan en el país para medir el logro de los estudiantes se limitan a ciertos campos del saber. PISA para el desarrollo, cuya toma concluyó en el mes de octubre de 2017, presentando resultados a inicios de 2019, levanta información sobre razonamiento matemático, comprensión lectora, pensamiento científico aplicado a las ciencias naturales y experimentales y solución de problemas entre una muestra de población de estudiantes con 15 años cumplidos; el LLECE evaluó, también de manera muestral, lectura, escritura y Matemática en 4° y 7° grados y, además, Ciencias Naturales en 7° Educación General Básica (EGB), en su última aplicación de 2015; el INEVAL, con sus pruebas Ser Estudiante, evaluó en su última aplicación las áreas de Lengua y Literatura, Matemática, Ciencias Naturales y Ciencias Sociales en 4°, 7° y 10° grados de EGB, en una muestra de estudiantes del ciclo 2015-2016; en las pruebas Ser Bachiller se abordan la aptitud abstracta y los dominios matemático, lingüístico, científico y social, estas pruebas son las únicas con carácter censal de las que se toman en el país, y tienen una periodicidad anual.

Dada la naturaleza del acto educativo, es probable que los estudiantes dominen una enorme cantidad de conocimientos o destrezas de cada una de las disciplinas que acabamos de enunciar que solo una prueba extremadamente larga podría abarcar, razón por la que las pruebas estandarizadas se diseñan a partir de un muestreo (POPHAM, 1999). En este sentido, el alineamiento se define como el nivel de correspondencia existente entre los estándares de contenido académico, el currículo oficial, la instrucción en el aula y la evaluación, que son los componentes curriculares que se busca alinear. Este concepto se relaciona con el de validez, considerando las evidencias teóricas y empíricas de un examen (MESSICK, 1989), ya que el grado de alineamiento de los ítems de un examen y los estándares deben considerarse como una condición necesaria para validar los exámenes.

El objetivo principal de las experiencias que han abordado un análisis del alineamiento es tener una perspectiva clara del nivel técnico de los instrumentos evaluativos en correspondencia con los estándares: identificar áreas problemáticas en las evaluaciones (LÓPEZ; WEBB; STANSFIELD, 2006; ROTHMAN, 2004), determinar la validez de los resultados y si son suficientemente representativos (HERMAN, 2004; PORTER, 2002; ROTHMAN, 2004; WISE, 2004) y establecer qué tipo de inferencias se pueden realizar sobre el proceso de aprendizaje que se deriva de los resultados (BEHUNIAK, 2004).

Case y Zucker (2005) identifican los tres métodos más usados para evaluar y documentar sistemáticamente el alineamiento entre estándares y evaluaciones: desarrollo secuencial, análisis de documentos y revisión de expertos.

El desarrollo secuencial implica la colaboración de expertos, educadores y público en general para aportar en la formulación de los estándares de aprendizaje; que sirven de base para plantear posteriormente la evaluación. Lo más destacado de este método es canalizar la correspondencia secuencial entre estándares y pruebas. Uno de los principales exponentes de este método es Paul La Marca (2001) quien reconoce la importancia del alineamiento como uno de los criterios para la evaluación, rendición de cuentas y toma de decisiones basadas en los logros académicos de los estudiantes. Su obra proporciona una visión general del concepto de alineamiento, a la vez que discute los problemas metodológicos que afectan su estudio y la interpretación del resultado.

El método basado en análisis de documentos examina, mediante plantillas de especificaciones que codifican su contenido y estructura, las normas y los formularios de las pruebas. Las matrices resultantes precisan la correspondencia entre los estándares, los contenidos, los procesos cognitivos y las preguntas de la prueba. Los niveles de complejidad cognitiva suelen apoyarse en la taxonomía de Bloom (BLOOM et al., 1956). Esta metodología es adecuada para estudios de alineamiento complejos y masificados como PISA, o comparar sistemas educativos entre países. Un grupo de expertos realiza el análisis y codificación.

Por último, el método basado en la revisión de expertos analiza el alineamiento entre estándares y evaluaciones después de que ambos han sido desarrollados. Existen métodos de alineamiento tradicionales y particulares, y entre estos últimos destacan el método *Survey* (SMITHSON; PORTER, 1994) y el método Webb (WEBB, 1997).

El método *Survey of enacted curriculum*, propuesto por Smithson y Porter (1994), propone analizar una prueba a partir de los estándares curriculares usando una matriz bidimensional de contenido y un cuestionario de prácticas instruccionales en el aula.

Norman Webb (1997) reflexiona sobre el concepto de alineamiento y examina los requerimientos para que las expectativas de estándares, contenidos y evaluaciones estén alineadas. Incluye en su trabajo: a) la importancia del alineamiento, b) los métodos: desarrollo secuencial, revisión por expertos y análisis de documentos; c) criterios específicos que incluyen el enfoque de contenido, articulación entre grados, edades, equidad, implicaciones pedagógicas y aplicabilidad del sistema. El autor concluye que el alineamiento de expectativas y evaluaciones es la clave del principio de reforma del sistema educativo.

El método Webb se ha empleado en diversos estudios de alineamiento. Su fiabilidad se atribuye entre otras cuestiones a que incorpora varias perspectivas de análisis reconocidas (HERMAN; WEBB; ZUNIGA, 2007; MARTONE; SIRECI, 2009) y a la revisión constante que realiza el autor de su planteamiento original (WEBB, 1997; 1999; 2002; 2006; 2007), poniendo de relieve las dificultades encontradas en cada experiencia.

En Latinoamérica, el método propuesto por Webb se ha aplicado para realizar el análisis del alineamiento entre los estándares académicos y la prueba Saber Matemáticas en Colombia (LÓPEZ, 2013), y para realizar el análisis del alineamiento de la demanda cognitiva entre los estándares educativos y la evaluación en Álgebra en Honduras (RAMOS; CASAS, 2016).

Criterios para el análisis del alineamiento según el método Webb

Norman Webb (2002) sugiere que el grado de alineamiento se mide de acuerdo con los siguientes criterios, cuyas descripciones se han ajustado a los elementos del currículo ecuatoriano:

- I. Concurrencia de categorías de contenido. Indica si los ítems de la prueba de evaluación incorporan las mismas categorías de contenido que las destrezas con criterios de desempeño (denominación de los contenidos del currículo ecuatoriano) del currículo, agrupadas por estándares de aprendizaje y bloques curriculares.
- II. Profundidad de conocimiento. Indica la relación entre la demanda cognitiva de los ítems en las evaluaciones y la expresada en las destrezas con criterios

- de desempeño del currículo, para lo que se establecen cuatro niveles de demanda cognitiva: memorístico, procesamiento, estratégico, extendido.
- III. Amplitud de conocimiento. Indica la cobertura del contenido de los ítems de las pruebas, de acuerdo con las destrezas con criterios de desempeño del currículo agrupadas por estándares de aprendizaje y bloques curriculares.
- IV. Balance de representación. Indica la distribución de los ítems de la prueba para cada destreza con criterios de desempeño del currículo representada, agrupadas por estándares de aprendizaje y bloques curriculares.
- V. Fuente de desafío. Además de los criterios anteriores, Webb (2002) indica que la resolución de un ítem puede presentar algún tipo de fuente de desafío, la cual obstaculiza la posibilidad de determinar el desempeño del estudiante al momento de resolver dicho ítem. Este análisis representa una observación cualitativa de carácter técnico o conceptual que hace un experto sobre un elemento del ítem de evaluación, es decir, identifica si algún ítem presenta en su resolución alguna problemática. Un ítem presenta un problema de fuente de desafío, sí:
1. Un estudiante que conoce el contenido del ítem puede escoger una respuesta incorrecta.
 2. Un estudiante que, aunque no conoce el contenido del ítem puede responderlo correctamente.

Método

Este estudio realiza un análisis cualitativo y cuantitativo a partir de la sistematización de las decisiones de codificación de dos paneles de 8 expertos con respecto a los criterios establecidos por Webb (2002) y detallados en el epígrafe anterior, para determinar el alineamiento entre las pruebas de evaluación —Ser Bachiller régimen Costa 2017 y 2018 (en Ecuador hay dos calendarios escolares, Costa y Sierra)— y los elementos curriculares —bloques curriculares, estándares de aprendizaje y destrezas con criterios de desempeño de las áreas de Lengua y Literatura, y Matemática en el bachillerato—.

Durante las diferentes sesiones de codificación, los paneles de expertos registraron sus decisiones individuales y consensuadas empleando instrumentos desarrollados y validados previamente en sendos estudios piloto (FIGUEROA CHÁVEZ; HERRERA PAVO, 2019) realizados para cada una de las áreas del currículo analizadas: (1) tablas de

conurrencia de categorías de contenido, (2) tablas de profundidad de conocimiento o demanda cognitiva, (3) tabla de análisis de la fuente de desafío.

Se calculó el coeficiente de correlación intraclase para las decisiones individuales relacionadas con la codificación de los ítems de evaluación de acuerdo con el criterio de demanda cognitiva o profundidad de conocimiento, siendo aceptable, mayor a 0,7, en todos los casos, menos en la prueba de Lengua y Literatura de 2017, que fue de 0,585.

Los elementos de análisis

El currículo de los niveles de la educación obligatoria de Ecuador se estructura en áreas para cada nivel y subnivel educativo, organizando los aprendizajes en bloques curriculares, cuyos contenidos son expresados en Destrezas con Criterios de Desempeño (DCD) (ECUADOR, 2016), que se distribuyen, a partir de los criterios de evaluación, en los estándares de aprendizaje. La estructura de las áreas curriculares en las que se enfoca el estudio se presenta en la Tabla 1, a continuación.

Tabla 1 - Desagregación de los componentes curriculares

Bloques	Estándares	# DCD	# DCD por bloque	# DCD por área
Lengua y cultura	E.LL.5.1.	2	4	27
	E.LL.5.2.	2		
Comunicación oral	E.LL.5.3.	4	4	
Lectura	E.LL.5.4.	5	6	
	E.LL.5.5.	1		
Escritura	E.LL.5.6.	8	8	
Literatura	E.LL.5.7.	3	5	
	E.LL.5.8.	2		
Álgebra y funciones	E.M.5.1.	8	78	130
	E.M.5.2.	11		
	E.M.5.3.	36		
	E.M.5.4.	9		
	E.M.5.5.	14		
Geometría y medida	E.M.5.6.	17	27	
	E.M.5.7.	6		
	E.M.5.8.	4		
Estadística y probabilidad	E.M.5.10.	15	25	
	E.M.5.11.	4		
	E.M.5.9.	6		

Fuente: Los autores (2019).

El instrumento Ser Bachiller está conformado por “dos componentes de potencial desarrollo académico y cognitivo: aptitudes y destrezas, integradas activamente en cada uno de los campos valorados. El instrumento Ser Bachiller evalúa la síntesis de los estándares educativos de sus cuatro dimensiones: Matemática, Lengua y

Literatura, Ciencias Naturales y Estudios Sociales, así como los campos de Razonamiento Verbal, Numérico y Abstracto." (INEVAL, 2017, p. 2). Esta prueba "es condición para graduarse de bachiller y aporta el 30% de la Nota final de Graduación del Bachiller." (INEVAL, 2017, p. 3).

La evaluación consta de 155 ítems, de los cuales 40 pertenecen al área de Lengua y Literatura y 40 a la de Matemática, todos están diseñados con respuesta cerrada y sistema de opción múltiple en la modalidad de aplicación cien por ciento.

Para el desarrollo del estudio se identificó que la prueba de 2017 contaba para ambas áreas con 41 ítems mientras que para 2018 sólo contaba con 40 ítems por área. Cada ítem tenía un peso y valor ponderado en 1, por lo que los ítems de todas las áreas tienen el mismo puntaje en la calificación global. En ese sentido, los ítems fueron analizados de forma única sin ponderaciones especiales al momento de la asignación de los índices del método Webb.

El registro y procesamiento de la información

De los criterios de alineamiento definidos para el análisis según el método Webb, la concurrencia de contenidos y la profundidad de conocimiento o demanda cognitiva requieren un proceso de codificación por parte de los expertos. Los criterios de amplitud de conocimiento y balance de representación de contenidos se construyen a partir de la información recogida para los dos primeros. Por último, la fuente de desafío recoge información sobre las posibles dificultades para la resolución de un ítem, como se ha comentado con anterioridad, a partir de un debate realizado por los expertos.

Concurrencia de las categorías de contenido

Cada experto revisó los ítems de las pruebas de evaluación con el objetivo de relacionarlos con las destrezas con criterios de desempeño que, a su juicio, evaluaban, y registró el resultado del análisis realizado de forma individual para, después, en una sesión de grupo, exponer sus decisiones y participar en un debate con el objeto de lograr consensos con respecto a las relaciones establecidas individualmente.

Se discutió la relación de cada ítem de evaluación con las destrezas del currículo en un máximo de dos rondas de debate, donde si no se alcanzaba un consenso por unanimidad se establecía por un mínimo de 5 votos, lo que correspondía a la mitad

más uno de los expertos de cada panel. Todas estas decisiones quedaron registradas en fichas individuales y de consenso.

Se analizó cada una de las decisiones individuales, agrupándolas por bloque curricular, estándar de aprendizaje y destreza con criterios de desempeño que cada experto asignó. Posteriormente se hizo un conteo por experto de todos los ítems asignados para cada destreza con criterios de desempeño. A continuación, se calculó la media aritmética para determinar cuántos ítems representaban las destrezas con criterios de desempeño por bloque curricular en promedio. Finalmente, se calculó la desviación estándar para determinar la variación en las decisiones individuales con respecto a la cantidad de ítems y destrezas con criterios de desempeño representadas por bloque curricular.

Con los resultados de estos cálculos, se determinó qué tipo de alineamiento mostraba el criterio de concurrencia de categorías de contenido de acuerdo con lo establecido en los estudios realizados por Webb (2002):

- Existe alineamiento cuando al menos se codifican, en promedio, seis ítems de evaluación en cada bloque curricular.
- El alineamiento es débil si los ítems de evaluación codificados en cada bloque curricular, en promedio, son cinco.
- No existe alineamiento si los ítems de evaluación codificados en cada bloque curricular, en promedio, son menos de cinco.

Profundidad de conocimiento o demanda cognitiva

A continuación, los expertos procedieron a determinar el nivel de profundidad de conocimiento o demanda cognitiva de cada uno de los ítems de evaluación y de cada una de las destrezas con criterios de desempeño del currículo relacionadas, primero de manera individual, y después en sesiones grupales, con la finalidad de alcanzar consensos.

Para el logro de consensos se procedió de la misma manera que en el caso anterior, se discutió el nivel de demanda cognitiva de cada ítem y de cada destreza del currículo en un máximo de dos rondas de debate, donde si no se alcanzaba un consenso por unanimidad se establecía por un mínimo de 5 votos, lo que correspondía a la mitad más uno de los expertos de cada panel. Todas estas decisiones quedaron registradas en fichas individuales y de consenso.

Con respecto a las decisiones individuales, se analizó si un experto consideraba que la demanda cognitiva de cada ítem era mayor, igual o menor que la demanda cognitiva de la destreza con criterios de desempeño con la que había determinado que existía una concurrencia de contenido.

Posteriormente se realizó un conteo por experto, para establecer, para cada uno de los bloques curriculares, cuántos ítems presentaban una demanda cognitiva mayor, igual o menor a las destrezas con criterios de desempeño asociadas.

Seguidamente, para cada bloque curricular, se calculó el promedio de las decisiones individuales de los ocho expertos y el porcentaje promedio total de ítems que presentaban una demanda cognitiva mayor, igual o menor a las destrezas con criterios de desempeño a las que habían sido asociados, así como su desviación estándar para determinar la variación en las decisiones de los expertos.

Por último, se calculó el coeficiente de correlación intraclase para el conjunto de las decisiones individuales tomadas por los expertos con respecto a la profundidad de conocimiento o demanda cognitiva de los ítems de cada una de las pruebas de evaluación, como medida de confiabilidad, considerando que los valores por encima de 0.7 para este coeficiente eran aceptables.

Con los resultados de estos cálculos, se determinó qué tipo de alineamiento mostraba el criterio de profundidad de conocimiento o demanda cognitiva de acuerdo con lo establecido en los estudios realizados por Webb (2002):

- Existe alineamiento cuando al menos, en promedio, un 50% o más de los ítems de evaluación consignados en cada bloque curricular presentan una demanda cognitiva igual o mayor que la destreza con criterios de desempeño con la que se determinó la existencia de concurrencia de contenido.
- El alineamiento es débil si no se alcanza este 50%, pero un 40% o más de los ítems de evaluación consignados en cada bloque curricular presentan, en promedio, una demanda cognitiva igual o mayor que la destreza con criterios de desempeño con la que se determinó la existencia de concurrencia de contenido.
- No existe alineamiento si menos del 40% de los ítems de evaluación consignados en cada bloque curricular presentan, en promedio, una demanda cognitiva igual o mayor que la destreza con criterios de

desempeño con la que se determinó la existencia de concurrencia de contenido.

Amplitud de conocimiento

Para establecer en qué medida una prueba cubría de manera adecuada el contenido presentado en cada uno de los bloques curriculares, se calculó cuántas destrezas con criterios de desempeño de las consignadas en cada uno de estos bloques concurrían en cuanto a contenido con los ítems de la prueba de evaluación.

Este análisis tomó en consideración las decisiones individuales de cada uno de los expertos, a partir de las cuales se calculó el porcentaje, en promedio, del total de destrezas que presentan concurrencia de contenido con los ítems de evaluación en cada bloque. Finalmente, se calculó la desviación estándar para determinar la variación de las decisiones de los expertos.

Con los resultados de estos cálculos, se determinó qué tipo de alineamiento mostraba el criterio de amplitud de conocimiento de acuerdo con lo establecido en los estudios realizados por Webb (2002):

- Existe alineamiento cuando, al menos, en promedio, un 50% o más de las destrezas con criterios de desempeño en cada bloque curricular concurren con algún ítem de la prueba de evaluación.
- El alineamiento es débil si no se alcanza este 50%, pero un 40% o más de más de las destrezas con criterios de desempeño en cada bloque curricular concurren con algún ítem de la prueba de evaluación.
- No existe alineamiento si menos del 40% de más de las destrezas con criterios de desempeño en cada bloque curricular concurren con algún ítem de la prueba de evaluación.

Balance de representación

Para determinar el índice de balance de representación en la prueba de evaluación de cada uno de los bloques curriculares, se calculó a partir de la concurrencia de contenidos cuántos ítems de una misma prueba evaluaban la misma destreza con criterios de desempeño en cada uno de los bloques curriculares. Esto nos permitió obtener un porcentaje promedio de ítems por destreza con criterios de desempeño en función del total de ítems de cada bloque curricular.

A continuación, se aplicó la ecuación para el cálculo del índice de balance de Webb para el conjunto de destrezas evaluadas por bloque curricular.

Finalmente se estimó la desviación estándar para determinar el rango de variación de las decisiones individuales.

Para el cálculo del índice de balance, la ecuación propuesta por Webb (2002) es:

$$1 - \left(\sum_{K=1} \left| \frac{1}{(DCD)} - \frac{I_k}{(H)} \right| \right) / 2$$

$K = 1$

Donde:

DCD = Número total de destrezas con criterios de desempeño representadas por bloque curricular.

I_k = Número de ítems correspondientes a la destreza con criterios de desempeño evaluada.

H = Número total de ítems de evaluación en cada bloque curricular.

Con los resultados de estos cálculos, se determinó qué tipo de alineamiento mostraba el criterio de balance de representación de acuerdo con lo establecido en los estudios realizados por Webb (2002):

- Existe alineamiento cuando obtenemos un índice de balance de, al menos, 0.7.
- El alineamiento es débil si el índice de balance es menor a 0.7, pero igual o mayor que 0.6.
- No existe alineamiento si el índice de balance está por debajo de 0.6.

Fuente de desafío

Para este criterio se analizaron cada uno de los ítems de forma individual para determinar si las evaluaciones presentaban algún tipo de inconsistencia, fueran estas de tipo técnico o conceptual.

Las observaciones técnicas pueden presentarse si:

- En el planteamiento o en las opciones de respuesta del ítem existen elementos que dificultan la comprensión de lo que se demanda saber o saber hacer para resolverlo.
- Se presenta algún tipo de información adicional que puede causar algún tipo de confusión.
- La redacción no es adecuada para la población a la que va dirigida la evaluación.

Para el caso de las observaciones de tipo conceptual:

- En el planteamiento del ítem existe algún tipo de error conceptual y en consecuencia no hay respuesta correcta en las opciones o éstas son ambiguas.
- El procedimiento implícito al que hace referencia el ítem no considera elementos conceptuales necesarios para resolverlo, por consiguiente, la respuesta señalada como correcta no lo es.

Al término de la revisión individual de cada experto, se sometieron a discusión las observaciones a cada ítem para llegar a un consenso que determinara si efectivamente existía una fuente de desafío y redactar la descripción de la dificultad presentada.

Este procedimiento se realizó haciendo una revisión de los parámetros psicométricos, proporcionados por el Instituto Nacional de Evaluación, para establecer si la observación de los expertos correspondía con una o más observaciones de alguno de estos parámetros.

Los parámetros utilizados fueron dificultad y discriminación, ambos utilizados en la teoría clásica de los testes (TCT) y la teoría de respuesta al ítem (IRT) para determinar la calidad técnica de los ítems. Adicionalmente a estos parámetros, se sometieron a revisión los datos resultantes del análisis de distractores, proporcionados en la misma tabla de parámetros psicométricos. Esta información proporciona los porcentajes de elección, por parte de los estudiantes de las opciones de respuesta del ítem. Se utilizaron estos porcentajes para identificar si las opciones de respuesta diferentes a la marcada como respuesta correcta presentaban algún tipo de fuente de desafío que pudiera motivar que los estudiantes la eligieran y así corroborar las observaciones realizadas por los expertos.

Resultados y discusión

Análisis de concurrencia en las categorías de contenido

En general, los ítems de las pruebas no incorporan todos los contenidos definidos en las destrezas con criterios de desempeño del currículo. Ningún ítem de evaluación se alinea por completo con alguna destreza con criterios de desempeño y, generalmente, el punto de alineamiento se encuentra en los contenidos conceptuales. Hay contenidos fácticos y de procedimiento simple que se priorizan y contenidos procedimentales complejos y actitudinales que se están dejando al margen de la evaluación. Además, varios ítems evalúan los mismos contenidos y en

la misma forma y profundidad. Por otro lado, hay ítems que, aunque pueden alinearse con algún bloque curricular, no pueden asociarse con ninguna destreza con criterios de desempeño, dado que corresponden a un nivel educativo inferior.

Tabla 2 - Concurrencia de categorías

Bloques			SB 2017			SB 2018		
			Ítems		A*	Ítems		A
Nombre	# Estándares	# DCD	\bar{X}	σ		\bar{X}	σ	
Lengua y cultura	2	4	0.0	0.0	No	0.2	0.7	No
Comunicación oral	1	4	5.0	5.4	Débil	3.4	2.6	No
Lectura	2	6	17.0	5.0	Sí	14.1	4.3	Sí
Escritura	1	8	16.9	5.8	Sí	19.7	4.0	Sí
Literatura	2	5	2.1	1.4	No	2.5	0.9	No
Algebra y funciones	5	78	28.2	1.8	Sí	27.8	1.4	Sí
Geometría y medida	3	27	5.4	1.6	Débil	4.2	1.4	No
Estadística y probabilidad	3	25	7.4	0.7	Sí	8.0	0.0	Sí

*A=Alineamiento

Fuente: Los autores (2019).

Para los bloques curriculares del área de Lengua y Literatura, la evaluación de 2017 presenta concurrencia con 10 de las 27 destrezas con criterios de desempeño, mientras que en 2018 con 11. En las dos evaluaciones, aproximadamente el 90% de los ítems evalúa los bloques de Escritura y Lectura; los temas con los que existe una mayor concurrencia son el uso de vocabulario preciso, el correcto significado de las palabras, aspectos formales de un texto y procesos para mejorar la comprensión autorregulada: uso de información previa, analogías, hipótesis, síntesis interna, deducción, entre otros procesos. Los bloques de Comunicación oral y Literatura se evalúan en forma somera y los contenidos comprendidos en estos son: registros del discurso oral, formatos de la comunicación y géneros literarios. El bloque Lengua y cultura no es evaluado.

En el área de Matemática, en 2017, se evalúan 17 de las 130 destrezas con criterios de desempeño del currículo de Matemática del bachillerato y, en 2018, 15.

La mayoría de los ítems de las pruebas de 2017 y 2018 evalúan contenidos del bloque de Algebra y Funciones, el 42%, siendo este el bloque curricular más extenso. En ambas evaluaciones los contenidos comúnmente evaluados en este bloque son: propiedades de orden de los números reales para resolver fórmulas de Física, Química y Biología, funciones reales, sucesiones numéricas reales, monótonas y definidas. En el bloque curricular de Geometría y medida, los contenidos evaluados abarcan las aplicaciones geométricas y físicas de vectores en el plano y problemas de

programación lineal. Además de la escasa concurrencia de contenido, en este bloque se evalúan parcialmente las destrezas con criterios de desempeño, pues las pruebas no abordan los contenidos actitudinales relacionados con el juicio de la validez de las soluciones en el contexto de un problema. En el bloque de Estadística y probabilidad las pruebas se centran en el concepto de probabilidad y la aplicación de métodos de conteo para la resolución de problemas.

Análisis de la profundidad de conocimiento o demanda cognitiva

En el área de Lengua y Literatura, el nivel de demanda cognitiva de las destrezas con criterios de desempeño es mayor que el de los ítems de las pruebas de evaluación. El débil alineamiento en profundidad de conocimiento responde, en gran medida, a que las destrezas con criterios de desempeño abarcan contenidos conceptuales, procedimentales y actitudinales, mientras que los ítems atienden sobre todo contenidos conceptuales. En algunos casos las destrezas con criterios de desempeño y los ítems abordan los mismos temas o contenidos, sin embargo, el saber hacer, el saber procedimental, siempre requiere una demanda cognitiva menor en los ítems que en las destrezas.

Los contenidos procedimentales que evalúan la prueba son: identificar (significado, palabra, paráfrasis, sinónimo, elemento metaforizado, tesis, idea principal, tipo de texto, inferencia), seleccionar, ordenar lógicamente y completar. Los contenidos procedimentales que exigen las destrezas con criterios de desempeño son: indagar, analizar causas, analizar críticamente, valorar, utilizar de manera selectiva y crítica, autorregular, consultar, recoger, comparar y organizar información, construir, defender, aplicar, usar, producir, expresar posturas, desarrollar un tema, recrear, experimentar.

Tabla 3 - Consistencia de profundidad de conocimiento, Ser Bachiller 2017

Bloques			Nivel del ítem							A
			#	% menor		% igual		% mayor		
Nombre	# E	# DCD		\bar{X}	σ	\bar{X}	σ	\bar{X}	σ	
Lengua y cultura	2	4	0	0.0	0.0	0.0	0.0	0.0	0.0	No
Comunicación oral	1	4	1	75.0	43.3	25.0	43.3	0.0	0.0	No
Lectura	2	6	17	65.6	23.0	25.0	15.3	9.4	14.0	No
Escritura	1	8	20	82.5	17.1	16.2	16.5	1.2	2.2	No
Literatura	2	5	3	70.8	30.9	29.2	30.9	0.0	0.0	No
Algebra y funciones	5	78	28	33.1	15.3	55.9	16.9	11.0	6.9	Sí
Geometría y medida	3	27	5	37.5	25.4	45.0	26.0	17.5	23.3	Sí
Estadística y probabilidad	3	25	8	31.2	20.8	50.0	26.5	18.7	22.5	Sí

Fuente: Los autores (2019).

Tabla 1 - Consistencia de profundidad de conocimiento, Ser Bachiller 2018

Bloques			Nivel del ítem							A
			#	% menor		% igual		% mayor		
Nombre	# E	# DCD		\bar{X}	σ	\bar{X}	σ	\bar{X}	σ	
Lengua y cultura	2	4	0	0.0	0.0	0.0	0.0	0.0	0.0	No
Comunicación oral	1	4	3	68.7	24.2	31.2	24.2	0.0	0.0	No
Lectura	2	6	14	61.6	21.1	27.7	14.0	10.7	12.9	No
Escritura	1	8	20	82.5	17.7	17.5	17.7	0.00	0.0	No
Literatura	2	5	3	50.0	37.3	37.5	35.1	12.5	33.1	Sí
Álgebra y funciones	5	78	28	34.6	20.7	53.7	17.7	11.8	9.3	Sí
Geometría y medida	3	27	4	53.1	36.3	28.1	23.2	18.7	34.8	Débil
Estadística y probabilidad	3	25	8	29.7	30.6	45.3	32.4	25.0	28.6	Sí

Fuente: Los autores (2019).

En cambio, en el área de Matemática, las pruebas muestran, en general, un nivel de demanda cognitiva similar al del currículo, a excepción de los ítems asignados al bloque curricular de Geometría y medida en el año 2018, cuyo alineamiento es débil. Este particular responde a que el saber hacer del 50% de las destrezas con criterios de desempeño evaluadas en este bloque, demandan mayor exigencia cognitiva que los ítems. El contenido procedimental que evalúa las pruebas en Geometría y medida consiste en resolver aplicaciones de modelo simple de línea de producción, mientras que las destrezas exigen resolver y plantear aplicaciones e interpretar y juzgar la validez de soluciones, es decir, saber hacer actitudinal y procedimental. En general, la demanda cognitiva de los ítems se enfoca en el nivel memorístico y de procedimientos, mientras que varias de las destrezas con criterios de desempeño consideradas buscan desarrollar pensamiento de procesamiento más complejo y abstracto.

No obstante, es preciso tener en cuenta que para este cálculo se excluyeron los 11 ítems de las pruebas de 2017 y 2018 que, si bien se podrían encuadrar en el bloque de Álgebra y funciones, no pueden relacionarse con ninguna destreza con criterios de desempeño del currículo de bachillerato.

Análisis de la amplitud de conocimiento

En el área de Lengua y Literatura, las dos pruebas de evaluación, 2017 y 2018, sólo cubren 4 de los 8 estándares de aprendizaje. El 50 % de los ítems se concentran en el estándar 6 y el 37% en el estándar 4, es decir, que hay una sobrevaloración de los estándares 6 y 4. Los estándares 3 y 8 son parcialmente cubiertos, y los estándares 1, 2, 5 y 7 no son tomados en consideración en la evaluación. Por ello, solo

encontramos una adecuada representación del bloque de Lectura, en ambas pruebas.

Tabla 2 - Correspondencia de rango de conocimiento, Ser Bachiller 2017

Bloques			# Ítems		Rango de destrezas				A
Nombre	# E	# DCD	\bar{X}	σ	#		% del total		
					\bar{X}	σ	\bar{X}	σ	
Lengua y cultura	2	4	0.0	0.0	0.0	0.0	0.0	0.0	No
Comunicación oral	1	4	5.0	5.4	1.6	1.3	40.6	32.9	Débil
Lectura	2	7	17.0	5.0	4.2	0.8	70.8	13.8	Sí
Escritura	1	8	16.9	5.8	4.0	1.1	50.0	14	Sí
Literatura	2	5	2.1	1.4	1.0	0.5	20.0	10.0	No
Algebra y funciones	5	78	28.2	1.8	10	1.9	12.8	2.4	No
Geometría y medida	3	27	5.3	1.5	2.7	0.4	10.0	1.6	No
Estadística y probabilidad	3	25	7.3	0.7	3.1	0.6	12.5	2.4	No

Fuente: Los autores (2019).

Tabla 6 - Correspondencia de rango de conocimiento, Ser Bachiller 2018

Bloques			# Ítems		Rango de destrezas				A
Nombre	# E	# DCD	\bar{X}	σ	#		% del total		
					\bar{X}	σ	\bar{X}	σ	
Lengua y cultura	2	4	0.2	0.6	0.1	0.3	3.1	8.3	No
Comunicación oral	1	4	3.4	2.6	1.4	1.1	34.4	27.8	No
Lectura	2	7	14.1	4.2	3.0	1.0	50.0	16.7	Sí
Escritura	1	8	19.7	4.0	4.1	0.8	51.6	9.8	Sí
Literatura	2	5	2.5	0.9	1.0	0.0	20.0	0.0	No
Algebra y funciones	5	78	27.7	1.3	9.6	1.2	12.3	1.5	No
Geometría y medida	3	27	4.2	1.3	2.1	0.3	7.8	1.2	No
Estadística y probabilidad	3	25	8.0	0.0	3.2	0.8	13.0	3.3	No

Fuente: Los autores (2019).

En cuanto a Matemática, los estándares 5, 7 y 11 no se representan en ninguna de las evaluaciones, mientras que los estándares 3 y 10 se encuentran sobrevalorados y el resto no logra una representación adecuada. El elevado número de destrezas con criterios de desempeño del área, especialmente en el bloque de Álgebra y funciones, dificulta una adecuada representación con una prueba de tan solo 40 ítems.

Balance de la representación

Para ambas áreas, las pruebas de evaluación de 2017 y de 2018 muestran un adecuado balance de representación. Solo el bloque curricular de Álgebra y funciones muestra un alineamiento débil. Por lo que la distribución de los ítems de evaluación en las destrezas con criterios de desempeño con las que concurren en las categorías de contenido es adecuada.

Tabla 7 - Correspondencia al balance de representación, Ser Bachiller 2017

Bloques			# Ítems		Índice de balance				A
Nombre	# E	# DCD	□	σ	% Ítems / total Ítems		Índice		
					□	σ	□	σ	
Lengua y cultura	2	4	0.0	0.0	0.0	0.0	0.0	0.0	-
Comunicación oral	1	4	5.0	5.4	12.2	13.2	0.9	0.1	Sí
Lectura	2	6	17.0	5.0	41.5	12.3	0.7	0.1	Sí
Escritura	1	8	16.9	5.8	41.2	14.2	0.7	0.1	Sí
Literatura	2	5	2.1	1.4	5.2	3.5	1.0	0.1	Sí
Algebra y funciones	5	78	28.2	1.8	68.9	4.5	0.6	0.0	Débil
Geometría y medida	3	27	5.4	1.6	13.1	3.8	0.8	0.1	Sí
Estadística y probabilidad	3	25	7.4	0.7	18.0	1.7	0.7	0.1	Sí

Fuente: Los autores (2019).

Tabla 8 - Correspondencia al balance de representación, Ser Bachiller 2018

Bloques			# Ítems		Índice de balance				A
Nombre	# E	# DCD	□	σ	% Ítems / total Ítems		Índice		
					□	σ	□	σ	
Lengua y cultura	2	4	0.2	0.7	0.6	1.6	1.0	0.0	Sí
Comunicación oral	1	4	3.4	2.6	8.4	6.5	0.9	0.1	Sí
Lectura	2	6	14.1	4.3	35.3	10.7	0.8	0.1	Sí
Escritura	1	8	19.7	4.0	49.4	10.0	0.8	0.1	Sí
Literatura	2	5	2.5	0.9	6.2	2.8	1.0	0.0	Sí
Algebra y funciones	5	78	27.7	1.4	69.4	3.5	0.6	0.0	Débil
Geometría y medida	3	27	4.2	1.4	10.6	3.5	0.9	0.1	Sí
Estadística y probabilidad	3	25	8.0	0.0	20.0	0.0	0.7	0.0	Sí

Fuente: Los autores (2019).

Fuente de desafío

Al contraponer las observaciones de los expertos con los datos psicométricos de los ítems Ser Bachiller, se determina que de manera general existen novedades consideradas fuente de desafío, es decir, que podrían ocasionar impacto en la selección de una respuesta incorrecta en la evaluación.

En el área de Lengua y Literatura, 17 ítems de ambas evaluaciones presentaron problemas de fuente de desafío.

Las observaciones de tipo conceptual estuvieron ligadas a la naturaleza de la lengua castellana. En seis ítems se pide ordenar con lógica semántica y sintáctica palabras o frases, sin embargo, este tipo de ítems pueden contener más de una respuesta correcta, ya que la lengua castellana no tiene una estructura rígida y una frase se puede ordenar de diversas formas manteniendo una estructura coherente y válida. Quizás este tipo de ítems no representa un desafío mayor para los sustentantes, sin embargo, al plantear que sólo existe una estructura semántica y sintáctica

correcta se contradice la propuesta del currículo, que aboga por un lenguaje amplio y flexible, basado en diversos contextos sociales. Un par de ítems solicita encontrar una consecuencia lógica, sin embargo, la respuesta correcta podría ser subjetiva, de acuerdo con el punto de vista del emisor o el receptor, por lo tanto, no habría una única opción de respuesta lógica y válida. Un ítem demanda identificar la idea principal en un texto, sin embargo, el texto propuesto es un poema, así que lo más adecuado sería pedir al sustentante que identifique el objeto lírico o la voz poética; identificar la idea principal estaría más acorde con un texto en prosa. Por último, un ítem pide identificar el tipo de texto, pero el texto al que se hace alusión tiene demasiadas características similares entre subgéneros narrativos, dificultando así la identificación del género correcto.

Las observaciones de tipo técnico son dos. Un par de ítems, que demandan recurrir a la lógica para su resolución, fueron considerados por los expertos como un tema para el área de Matemática. Cinco ítems presentaron órdenes escuetas, sin contexto, por ejemplo: "complete el texto" donde no se indica al sustentante cuál es el contenido o demanda que evalúa la pregunta. Quizás ese tipo de ítems no representa dificultad para el sustentante, más bien se completa de forma inadvertida porque al carecer de contexto solo se requiere de la memoria, lo cual sugiere una demanda cognitiva básica.

En cuanto al área de Matemática, se identificaron 3 ítems con fuente de desafío, 1 ítem en la prueba de 2017 y 2 ítems en la de 2018. Los enunciados de estos ítems exigen relectura, demandando un elevado nivel de comprensión lectora y su resolución requiere más tiempo del promedio establecido para estas pruebas, además, las opciones de respuesta que sirven como distractores pueden causar confusión.

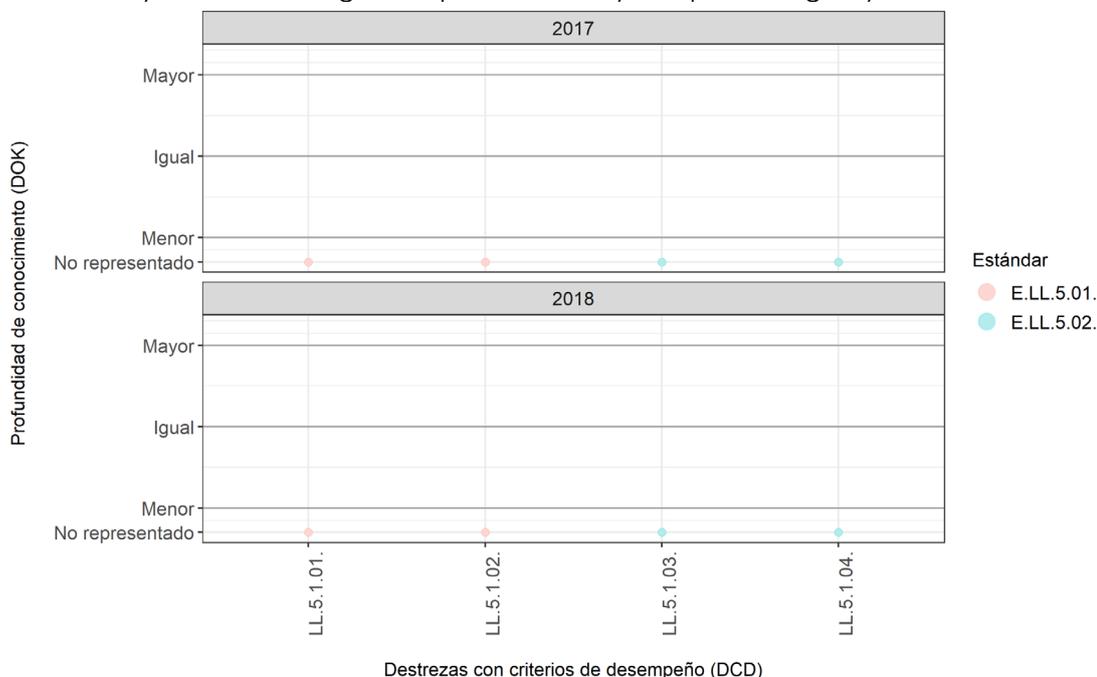
Ajuste de los resultados en función de los consensos

Los ejercicios de consenso realizados tras la puesta en común del trabajo individual de codificación, cuyos resultados se muestran en los gráficos de la figura 1 a la figura 7, promovieron un profundo debate con respecto a las codificaciones individuales que permitió la realización de ajustes a los resultados presentados.

En el caso de la concurrencia de contenidos estos debates no promovieron cambios relevantes en el alineamiento desde la perspectiva de los bloques curriculares, aunque sí hubo cambios en las destrezas con criterios de desempeño

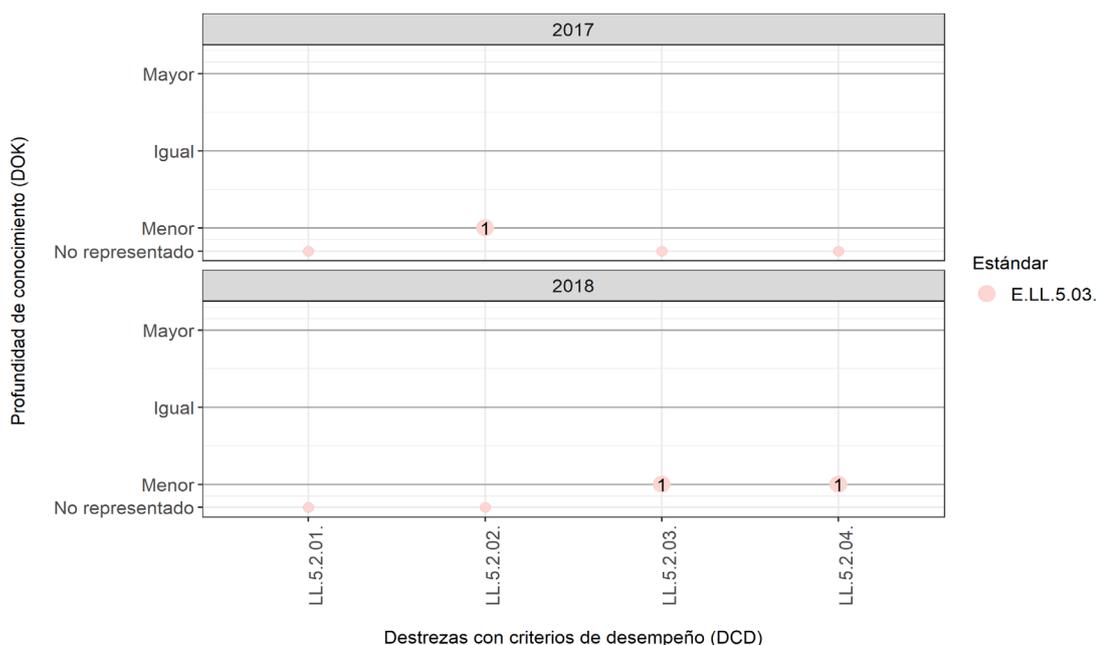
asignadas dentro de éstos, incluso considerando algunas que no habían sido tenidas en cuenta en la fase de codificación individual.

Figura 1 - Distribución de los ítems de la evaluación por destrezas con criterios de desempeño y demanda cognitiva, por estándar y bloque – Lengua y cultura



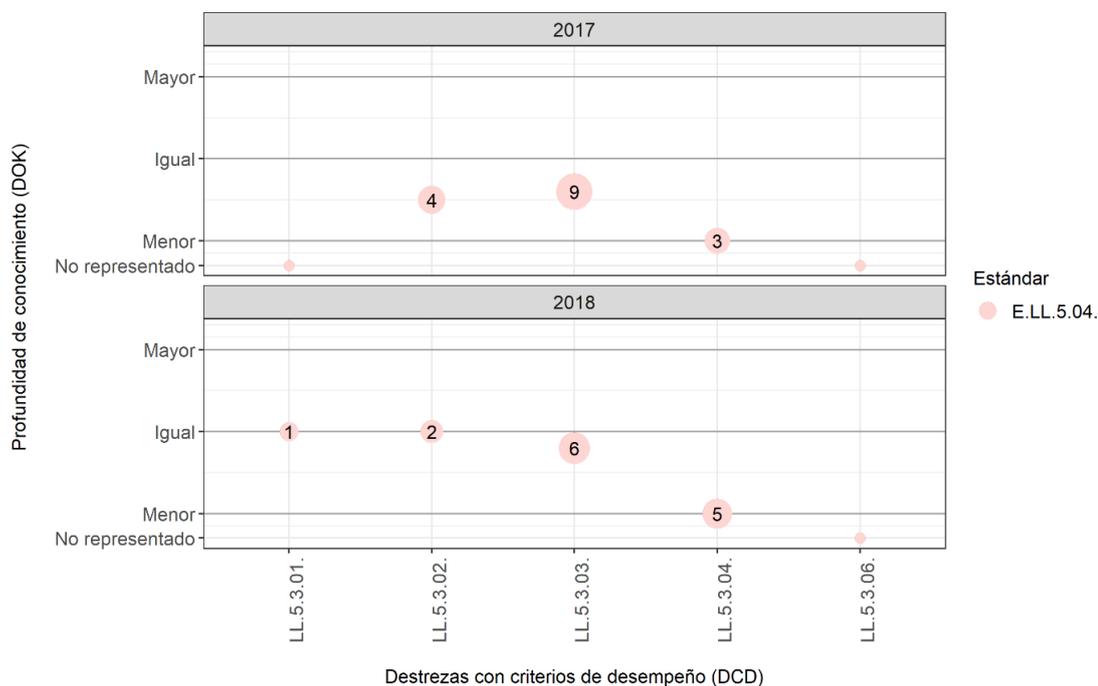
Fuente: Los autores (2019).

Figura 2 - Distribución de los ítems de la evaluación por destrezas con criterios de desempeño y demanda cognitiva, por estándar y bloque – Comunicación oral



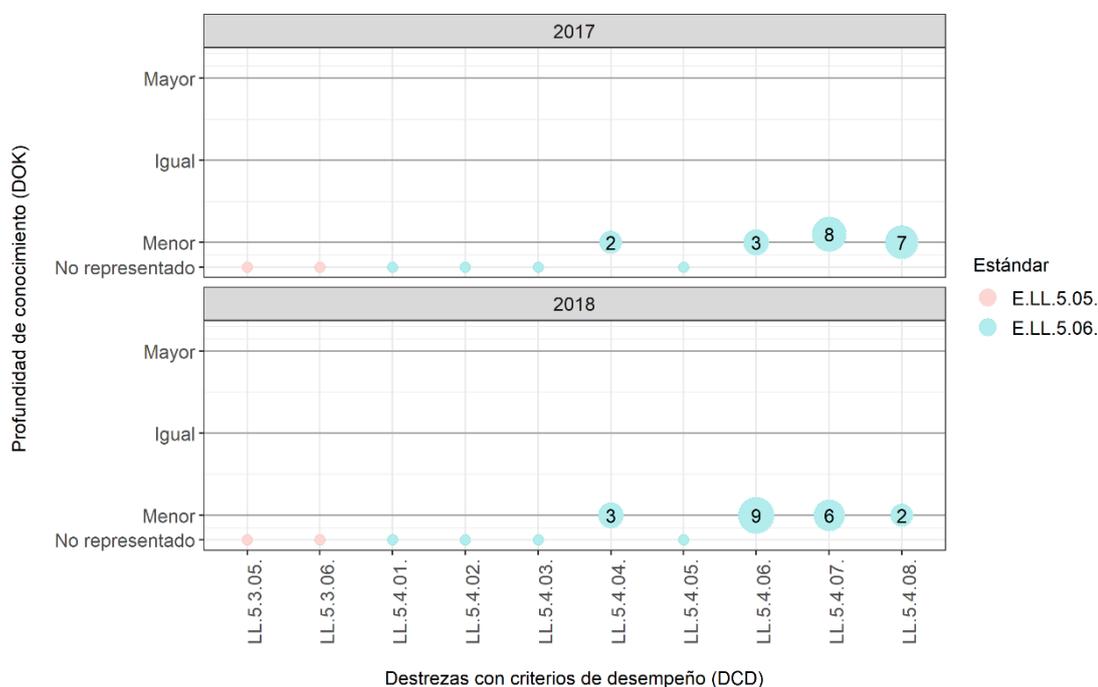
Fuente: Los autores (2019).

Figura 3 - Distribución de los ítems de la evaluación por destrezas con criterios de desempeño y demanda cognitiva, por estándar y bloque – Lectura



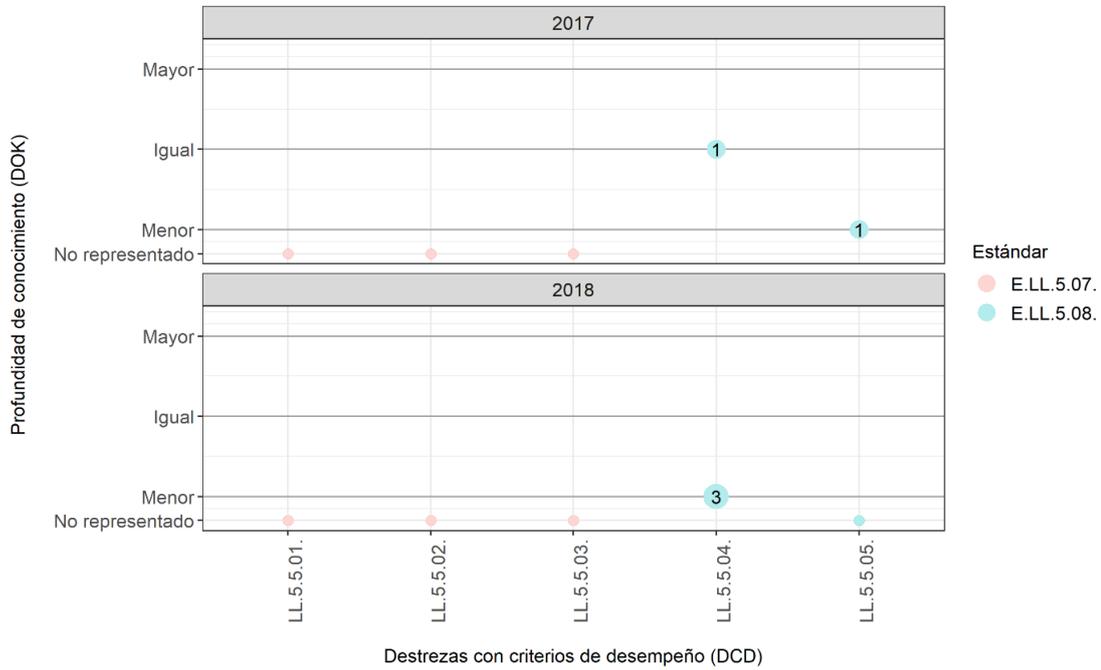
Fuente: Los autores (2019).

Figura 4 - Distribución de los ítems de la evaluación por destrezas con criterios de desempeño y demanda cognitiva, por estándar y bloque – Escritura



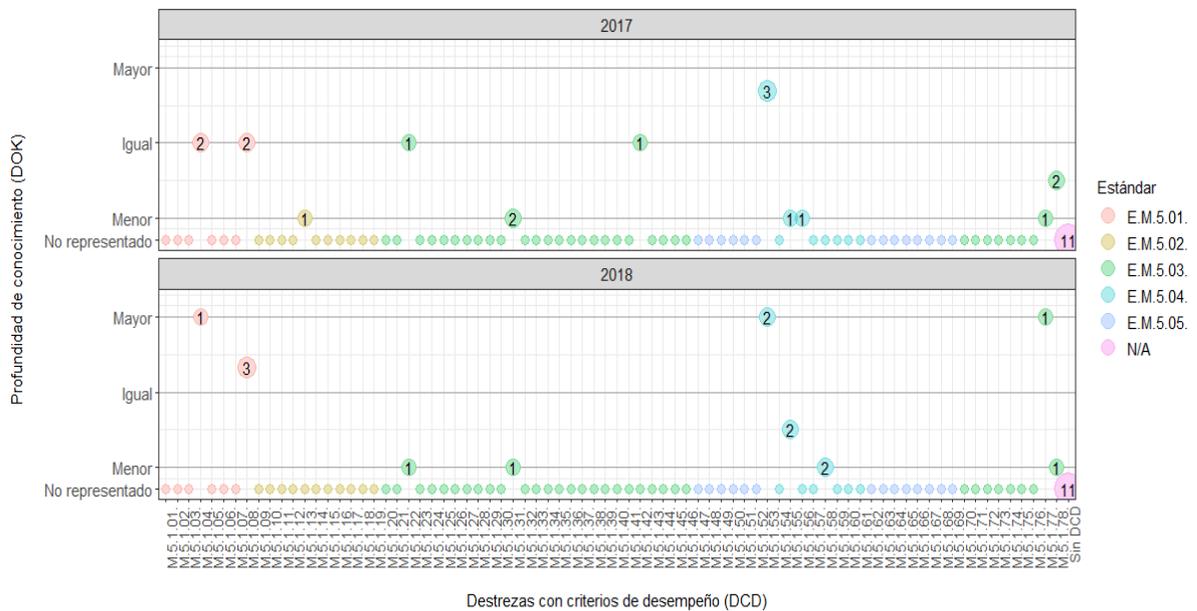
Fuente: Los autores (2019).

Figura 5 - Distribución de los ítems de la evaluación por destrezas con criterios de desempeño y demanda cognitiva, por estándar y bloque – Literatura



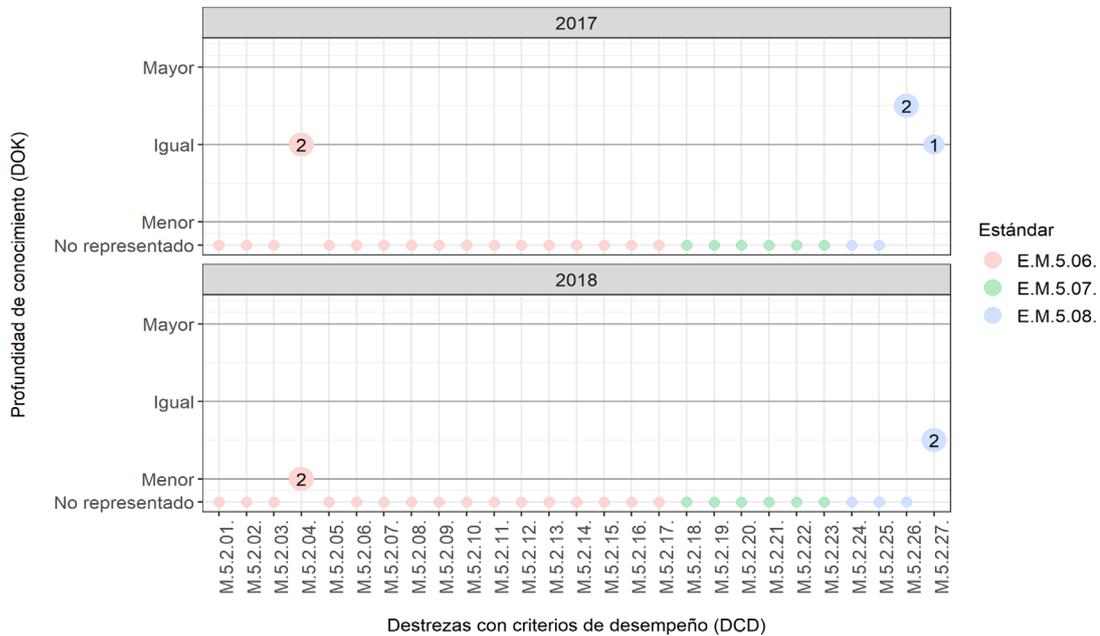
Fuente: Los autores (2019).

Figura 6 - Distribución de los ítems de la evaluación por destrezas con criterios de desempeño y demanda cognitiva, por estándar y bloque – Álgebra y funciones



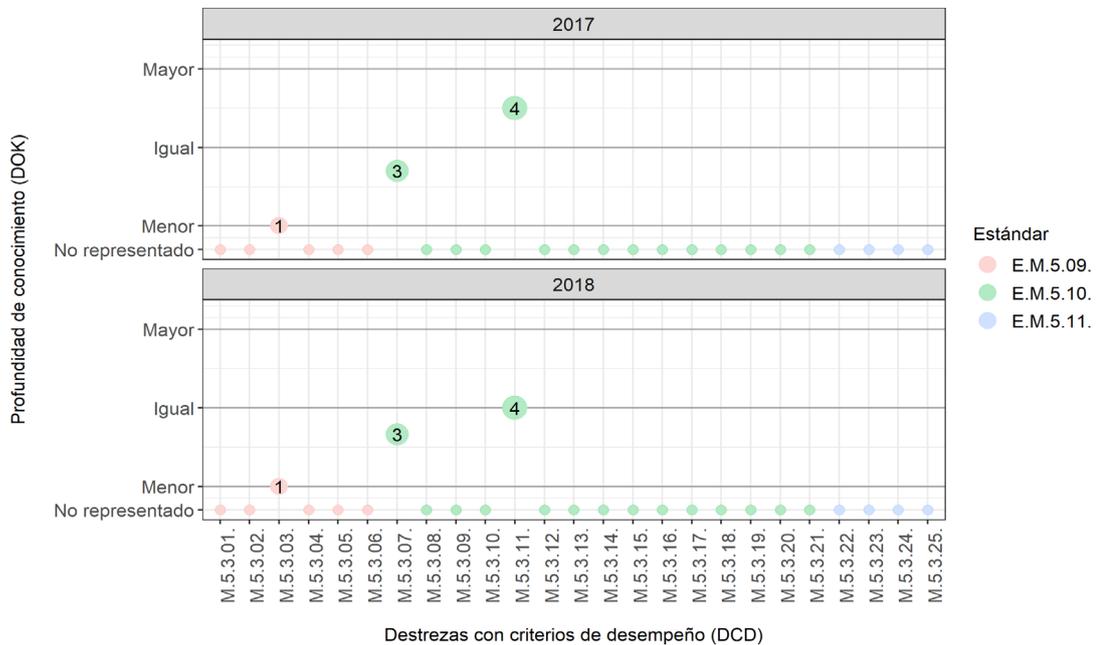
Fuente: Los autores (2019).

Figura 7 - Distribución de los ítems de la evaluación por destrezas con criterios de desempeño y demanda cognitiva, por estándar y bloque – Geometría y medida



Fuente: Los autores (2019).

Figura 8 - Distribución de los ítems de la evaluación por destrezas con criterios de desempeño y demanda cognitiva, por estándar y bloque – Estadística y probabilidad



Fuente: Los autores (2019).

En cuanto a la profundidad de conocimiento, los consensos alcanzados sirvieron para confirmar la ausencia de alineamiento en todos los bloques curriculares, para ambas pruebas, salvo en el bloque de Literatura que había obtenido un resultado positivo para la prueba de 2018 en la codificación individual, que ahora se traslada

a la de 2017. Si bien, en lo que respecta a Matemática, podemos ver cómo el debate establecido en torno al bloque de Álgebra y funciones hace que se reconsidere su alineamiento con las dos pruebas, después de los consensos, al considerar que los 11 ítems no asignados a ninguna destreza presentan una demanda cognitiva baja con respecto al currículo del nivel de bachillerato. Del mismo modo, se establece la falta de alineamiento del bloque de Geometría y medida para la prueba de 2018, para la que inicialmente se calculó un alineamiento débil.

En cuanto a la amplitud de conocimiento, la reasignación de algunos ítems a otras destrezas con criterios de desempeño a partir de los debates supuso cambios en el bloque de Comunicación oral, que resultan alineados en el caso de la prueba de 2018, y en el de Escritura, donde no se evidencia el alineamiento que se constataba según los resultados del ejercicio individual. En cualquier caso, el índice de balance permanece sin variaciones en todos los casos, excepto en el bloque de Lengua y cultura, donde, después del consenso, desaparece el único ítem asignado en la fase de codificación individual.

Por otro lado, se realizó un ejercicio comparativo de las pruebas de 2017 y 2018 que ayudó a establecer su similitud para los diversos parámetros, evidenciando que en estos dos años se evaluaron contenidos muy similares.

Conclusiones

El análisis de alineamiento muestra, para las áreas en las que se ha trabajado, que el esfuerzo por evaluar el currículo no es completamente adecuado, sería preciso optimizar la evaluación con el objetivo de aproximarla al trabajo prescrito a los estudiantes en el aula. A continuación, en la Tabla 9, se ofrece el resumen del análisis con los ajustes realizados en la fase de consensos.

Tabla 9 - Resumen del análisis de alineamiento

Bloques	SB 2017				SB 2018			
	Concurrencia	Consistencia	Rango	Balance	Concurrencia	Consistencia	Rango	Balance
Lengua y cultura	No	No	No	No	No	No	No	No
Comunicación oral	No	No	No	Sí	No	No	Sí	Sí
Lectura	Sí	No	Sí	Sí	Sí	No	Sí	Sí
Escritura	Sí	No	Sí	Sí	Sí	No	Sí	Sí
Literatura	No	Sí	No	Sí	No	No	No	Sí
Algebra y funciones	Sí	No	No	Débil	Sí	No	No	Débil
Geometría y medida	Débil	Sí	No	Sí	No	No	No	Sí
Estadística y probabilidad	Sí	Sí	No	Sí	Sí	Sí	No	Sí

Fuente: Los autores (2019).

En el área de Lengua y Literatura, la prueba debe afrontar importantes cambios si quiere estar mejor alineada con el currículo. Por un lado, la demanda cognitiva de la prueba es inferior a lo que el currículo requiere de los estudiantes, simplificando las demandas de las destrezas con criterios de desempeño, hecho que se revela tanto en la prueba de 2017, como en la de 2018, y que se sustenta en el análisis cualitativo realizado. Por otro lado, el nivel de concurrencia de las categorías de contenido es muy pobre, dejando fuera de la evaluación bloques curriculares fundamentales, como el de Literatura y el de Lengua y cultura, además del de Comunicación oral; este último, por las características de la prueba estandarizada. Por último, la distribución de los ítems de evaluación en los diferentes bloques curriculares solo cumple en los bloques de Lectura y Escritura, y en estos casos apuntando siempre al mismo conjunto de destrezas con criterios de desempeño.

La evaluación Ser Bachiller, en el área de Lengua y Literatura podría replantearse algunos aspectos para ser más pertinente: utilizar textos genuinos más amplios, pues los que se presentan son muy breves, lejanos de los estudiantes y del mismo tipo; evitar la sobre-evaluación de un mismo contenido y de la misma forma; construir ítems con indicaciones más específicas de lo que se espera que hagan los sustentantes teniendo en cuenta los postulados, ideología o modelos pedagógicos sobre los que está construido el currículo. Finalmente, considerar otros tipos de ítem o formas de evaluar más pertinentes con el área.

En el área de Matemática el análisis de alineamiento muestra un problema en el rango de destrezas con criterios de desempeño cubiertas, debido a la cantidad de destrezas que el currículo del bachillerato dispone (130) y el número de ítems de la prueba de evaluación (40). Este resultado demandaría un incremento de los ítems para la evaluación del área, y una mejor distribución de éstos en los bloques curriculares, ya que el bloque de Geometría y medida es el que tiene un tratamiento más pobre en las pruebas de evaluación, en cuanto a concurrencia. Por otro lado, el bloque de Álgebra y funciones requiere que todos los ítems de evaluación estén alineados a las destrezas con criterios de desempeño del nivel de bachillerato, para que exista un alineamiento adecuado en cuanto a demanda cognitiva. Finalmente, el análisis revela que quedan fuera de la prueba una cantidad considerable de contenidos altamente significativos desde el punto de vista curricular.

Estudios similares (LÓPEZ, 2013; LÓPEZ; WEBB; STANSFIELD, 2006; RAMOS; CASAS, 2016) concluyen que las pruebas estandarizadas deben mejorar su alineamiento en

cuanto a demanda cognitiva y representación de los estándares de aprendizaje y el currículo, expresando su inquietud por los límites de las pruebas estandarizadas y la necesidad de considerar la evaluación que se realiza en el aula como complemento para una adecuada interpretación de sus resultados.

En nuestro caso, el débil o inadecuado alineamiento de los estándares, el currículo y la prueba Ser Bachiller, de hecho, puede incidir en el aula, ya que autoridades y docentes, pueden tomar decisiones, llevados de la necesidad de mantener el prestigio institucional que supone un buen resultado en las pruebas de evaluación externa, que minimicen el desarrollo del currículo, dejen de lado contenidos importantes desde la perspectiva curricular y se orienten a preparar a los estudiantes exclusivamente para la evaluación externa.

Esta problemática se acentúa por el hecho de que las pruebas, de un año a otro, evalúan contenidos curriculares muy similares, obligando a concentrar los esfuerzos en la preparación de los exámenes de evaluación externa en idénticas parcelas del currículo, promoviendo que aquellas que quedan excluidas de las pruebas sean también desatendidas en el aula, ya que esta prueba, como se comentó anteriormente supone un 30% de la nota de graduación de los bachilleres.

En este sentido, sería preciso abordar una reforma de las pruebas orientada a atender de mejor manera los contenidos de cada uno de los bloques curriculares en las dos áreas curriculares analizadas. En el caso de Lengua y Literatura, es preciso adaptar la naturaleza y demanda cognitiva de los ítems de las pruebas de evaluación a los requerimientos del área. En el caso de Matemática sería preciso considerar la inclusión de un número mayor de ítems para mejorar el rango de representación de los contenidos en cada uno de los bloques curriculares, así como ajustar la demanda cognitiva de los ítems al nivel de bachillerato.

Referencias

- BEHUNIAK, P. Education assessment in an era of accountability. In: WALL, J. E.; WALZ, G. R. (org.). *Measuring up: assessment issues for teachers, counselors, and administrators*. Washington, DC: ERIC Counseling and Student Services, 2004. p. 335–347. Disponible en: <https://files.eric.ed.gov/fulltext/ED480379.pdf>. Acceso en: 12 oct. 2019.
- BLOOM, B. S. et al. *Taxonomy of educational objectives: the classification of educational goals*. Londres: Longmans, 1956. Disponible en: [https://www.uky.edu/~rsand1/china2018/texts/Bloom et al -Taxonomy of Educational Objectives.pdf](https://www.uky.edu/~rsand1/china2018/texts/Bloom%20et%20al%20-Taxonomy%20of%20Educational%20Objectives.pdf). Acceso en: 12 oct. 2019.
- CASE, B.; ZUCKER, S. *Methodologies for alignment of standards and assessments*. In: CONFERENCE ON ALIGNMENT OF ASSESSMENTS AND INSTRUCTION, 2005, Beijing. *Apresentation* [...]. Beijing: [s. n.], 2005. Disponible en: https://images.pearsonassessments.com/images/tmrs/tmrs_rg/AlignmentMethodologies.pdf?WT.mc_id=TMRS_Methodologies_for_Alignment. Acceso en: 14 mar. 2018.
- FIGUEROA CHÁVEZ, J. F.; HERRERA PAVO, M. A. Validación del método y los instrumentos del modelo Webb para el análisis del alineamiento entre el currículo de Matemática de bachillerato y la prueba Ser Bachiller. *593 Digital Publisher CEIT*, Ecuador, v. 4, n. 4, p. 4–18, 2 jul. 2019. Disponible en: http://www.593dp.com/index.php/593_Digital_Publisher/article/view/91. Acceso en: 12 oct. 2019.
- HERMAN, J. L. The effects of testing on instruction. In: FUHRMAN, S.; ELMORE, R. F. (org.). *Redesigning accountability systems for education*. Nueva York: Teachers College Press, 2004. p. 141–166. Disponible en: <https://www.tcpres.com/redesigning-accountability-systems-for-education-9780807744253>. Acceso en: 12 oct. 2019.
- HERMAN, J. L.; WEBB, N. M.; ZUNIGA, S. A. Measurement issues in the alignment of standards and assessments. *Applied Measurement in Education*, [S. l.], v. 20, n. 1, p. 101–126, 5 ene. 2007. Disponible en: <https://www.tandfonline.com/doi/full/10.1080/08957340709336732>. Acceso en: 12 oct. 2019.
- INEVAL. *Ser Bachiller 2017: ficha técnica y conceptual*. Quito: INEVAL, 2017. Disponible en: http://www.evaluacion.gob.ec/evaluaciones/wp-content/uploads/2017/07/Ineval_fichaSBAC17_20170224.pdf. Acceso en: 12 oct. 2019.
- LA MARCA, P. M. Alignment of standards and assessments as an accountability criterion. *Practical Assessment, Research & Evaluation*, [S. l.], v. 7, n. 21, p. 1–4, 2001. Disponible en: <https://pareonline.net/getvn.asp?v=7&n=21>. Acceso en: 12 oct. 2019.
- LINN, R. Assessments and accountability. *Educational Researcher*, Washington, v. 29, n. 2, p. 4–16, 1 mar. 2000. Disponible en: <http://journals.sagepub.com/doi/10.3102/0013189X029002004>. Acceso en: 20 nov. 2017.
- LÓPEZ, A. A. Alineación entre las evaluaciones externas y los estándares académicos: el caso de la prueba saber de matemáticas en Colombia. RELIEVE -

Revista *Electrónica de Investigación y Evaluación Educativa*, España, v. 19, n. 2, 9 ene. 2013. Disponible en: <http://ojs.uv.es/index.php/RELIEVE/article/view/3024>. Acceso en: 23 abr. 2019.

LÓPEZ, A. A.; WEBB, N. L.; STANSFIELD, C. W. *Alignment of the New Mexico language arts frameworks and the spanish reading standards-based assessments*. Santa Fe: [s.n.], 2006. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.577.8428>. Acceso en: 12 oct. 2019.

MARTONE, A.; SIRECI, S. G. Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research*, Washington, v. 79, n. 4, p. 1332–1361, 1 dic. 2009. Disponible en: <http://journals.sagepub.com/doi/10.3102/0034654309341375>. Acceso en: 12 oct. 2019.

MESSICK, S. Meaning and values in test validation: the science and ethics of assessment. *Educational Researcher*, Washington, v. 18, n. 2, p. 5–11, 1 mar. 1989. Disponible en: <http://journals.sagepub.com/doi/10.3102/0013189X018002005>. Acceso en: 12 oct. 2019.

ECUADOR. Ministerio de Educación. *Currículo de los niveles de educación obligatoria*. Quito: Ministerio de Educación, 2016. Disponible en: <https://educacion.gob.ec/wp-content/uploads/downloads/2016/08/Curriculov2.pdf>. Acceso en: 16 oct. 2018.

NORMAND, R. Les comparaisons internationales de résultats: problèmes épistémologiques et questions de justice. *Education et Sociétés*, Louvain, v. 12, n. 2, p. 73–89, 2003. Disponible en: <http://www.cairn.info/revue-education-et-societes-2003-2-page-73.htm>. Acceso en: 12 oct. 2019.

PÉREZ GÓMEZ, A. I. *Educarse en la era digital*. Madrid: Ediciones Morata, 2012.

PÉREZ GÓMEZ, A. I.; SOTO GÓMEZ, E. Luces y sombras de PISA: sentido educativo de las evaluaciones externas. *Cultura y Educación*, Madrid, v. 23, n. 2, p. 171–182, 23 ene. 2011. Disponible en: <https://www.tandfonline.com/doi/full/10.1174/113564011795944758>. Acceso en: 12 oct. 2019.

POPHAM, W. J. Why standardized tests don't measure educational quality: educational leadership. *Educational Leadership*, Washington, v. 56, n. 6, p. 8–15, 1999. Disponible en: <http://www.ascd.org/publications/educational-leadership/mar99/vol56/num06/Why-Standardized-Tests-Don't-Measure-Educational-Quality.aspx>. Acceso en: 12 oct. 2019.

PORTER, A. C. Measuring the content of instruction: uses in research and practice. *Educational Researcher*, Washington, v. 31, n. 7, p. 3–14, 1 oct. 2002. Disponible en: <http://journals.sagepub.com/doi/10.3102/0013189X031007003>. Acceso en: 12 oct. 2019.

RAMOS, L.; CASAS, L. Demanda cognitiva en estándares educativos y evaluación en álgebra. 2016, Málaga: SEIEM, 2016. Disponible en: <https://riuma.uma.es/xmlui/handle/10630/12564>. Acceso en: 12 oct. 2019.

ROTHMAN, R. Benchmarking and alignment of state standards and assessments. In: FUHRMAN, S.; ELMORE, R. F. (org.). *Redesigning accountability systems for education*. Nueva York: Teachers College Press, 2004. p. 96–114. Disponible en: <https://www.tcpres.com/redesigning-accountability-systems-for-education-9780807744253>. Acceso en: 12 oct. 2019.

SMITHSON, J. L.; PORTER, A. C. *Measuring classroom practice: lessons learned from efforts to describe the enacted curriculum: the reform up close study*. New Brunswick, NJ: Carriage House at the Eagleton Institute of Politics, 1994. Disponible en: <https://eric.ed.gov/?id=ED377563>. Acceso en: 12 oct. 2019.

WEBB, N. L. *Report [of] alignment analysis of mathematics standards and assessments*. [Wisconsin: s. n.], 2006. Disponible en: <https://dpi.wi.gov/sites/default/files/imce/assessment/pdf/mathsummary06.pdf>. Acceso en: 12 oct. 2019.

WEBB, N. L. *Alignment of science and mathematics standards and assessments in four states*. Washington, DC: National Institute for Science Education (NISE), 1999. Disponible en: <https://eric.ed.gov/?id=ED440852>. Acceso en: 12 oct. 2019.

WEBB, N. L. An analysis of the alignment between mathematics standards and assessments for three states. In: AMERICAN EDUCATIONAL RESEARCH ASSOCIATION ANNUAL MEETING, 2002. *Apresentation* [...]. Louisiana: [s. n.], 2002. Disponible en: <http://facstaff.wceruw.org/normw/AERA 2002/Alignment Analysis three states Math Final 31502.pdf>. Acceso en: 12 oct. 2019.

WEBB, N. L. *Criteria for alignment of expectations and assessments in mathematics and science education*. Washington, DC: National Institute for Science Education (NISE), 1997. Disponible en: <https://eric.ed.gov/?id=ED414305>. Acceso en: 12 oct. 2019.

WEBB, N. L. Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, [S. l.], v. 20, n. 1, p. 7–25, 5 ene. 2007. Disponible en: <https://www.tandfonline.com/doi/full/10.1080/08957340709336728>. Acceso en: 12 oct. 2019.

WISE, L. *Vertically-articulated content standards*. [s. l.], 2004. Interactive Lecture Series. Disponible en: http://www.nciea.org/sites/default/files/publications/RILS_LW04.pdf. Acceso en: 12 oct. 2019.

Agradecimientos

El presente trabajo de investigación fue realizado gracias al financiamiento de la Organización de Estados Iberoamericanos del Ecuador y al apoyo de la Universidad Andina Simón Bolívar, sede Ecuador. No hubiera sido posible realizar este trabajo sin la colaboración del Ministerio de Educación del Ecuador y del Instituto Nacional de Evaluación del Ecuador (INEVAL).