

1

**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**

Fernanda Figueroa 1, Miguel Herrera Pavo 2
1,2 Universidad Andina Simón Bolívar, 1 fercitafigueroa@hotmail.com ,
2 miguel.herrera.p@uasb.edu.ec

Fecha de presentación: 10 de marzo de 2019

Fecha de aceptación: 27 de abril de 2019

DOI: <https://doi.org/10.33386/593dp.2019.4.91>

RESUMEN

El alineamiento puede definirse como la medida en la que las finalidades educativas expresadas en el proyecto curricular y las evaluaciones estandarizadas externas se articulan para facilitar el desarrollo esperado de los y las estudiantes y servir de guía a los y las docentes. Existen diferentes modelos para el análisis del alineamiento, pero el modelo Webb es uno de los preferidos, debido a que ha sido aplicado en numerosas ocasiones y está ampliamente documentado. Este estudio valida el método y los instrumentos propuestos por el modelo Webb para su posterior aplicación en una investigación de alineación, siendo la primera vez que en el Ecuador se realiza un trabajo de este tipo. La validación del modelo Webb se realizó en diez fases, comenzando por la conformación y capacitación de un panel de 4 expertos, y prosiguiendo con el análisis, en sesiones individuales y de consenso, de cada una de las cinco categorías propuestas por Webb. El estudio supuso una dedicación de 25 horas para los expertos. El trabajo realizado por el panel de expertos permitió realizar ajustes para la mejora de la metodología y de los instrumentos de análisis. El estudio revela que la aplicación del modelo Webb en el contexto ecuatoriano es posible y se ve facilitada por la estructura del currículo 2016 y su articulación con los estándares de aprendizaje.

Palabras clave: alineación, currículo, evaluación estandarizada, Matemática

ABSTRACT

Alignment can be defined as the extent to which the educational goals expressed in the curricular project and the external standardized evaluations are articulated to facilitate the expected development of the students and serve as a guide for the teachers. There are different models for the analysis of the alignment, but the Webb model is one of the preferred ones, because it has been applied in many occasions and is widely documented. This study validates the method and instruments proposed by the Webb model for its later application in an alignment research, being the first time that a work of this type is carried out in Ecuador. The validation of the Webb model was developed in ten phases, starting with the formation and training of a panel of 4 experts, and continuing with the analysis, in individual and consensus sessions, of each of the five categories of analysis proposed by Webb. This work involved a dedication of 25 hours for the experts. The work carried out by the panel of experts allowed adjustments to be made to improve the methodology and the analysis tools. The study reveals that the application of the Webb model in the Ecuadorian context is feasible and is facilitated by the structure of the 2016 curriculum and its articulation with the learning standards.

Keywords: alignment, curriculum, standardized test, Mathematics

Introducción

La evaluación educativa estandarizada externa es reciente, pero suscita un gran interés a nivel mundial. En Latinoamérica su aplicación se extiende a partir de 1990, como resultado del desarrollo de reformas educativas regionales que han provocado cambios en ciertos principios tradicionales de la práctica educativa (Martinic, 2010). En este sentido, en varios países se crean instituciones dedicadas a la evaluación externa cuyo objetivo público es promover la educación de calidad, implementando evaluaciones externas para dar cuenta de los logros, comprender los problemas y realizar los cambios necesarios para mejorar los procesos y resultados educativos.

La preocupación de las reformas educativas de los últimos años por la calidad y equidad en la educación ha ampliado los aspectos de la realidad que son objeto de estas evaluaciones (Martinic, 2010). El Ecuador no se ha quedado al margen de estos procesos y en 2012 se creó el Instituto Nacional de Evaluación Educativa, INEVAL. A partir de ese año, el país participa en las evaluaciones estandarizadas internacionales: LLECE y PISA, y nacionales: las pruebas Ser Estudiante y Ser Bachiller, influyendo esta última en el 30% de calificación de graduación del bachillerato y siendo determinante para el acceso a la universidad pública (INEVAL, 2017, p.3).

Las evaluaciones externas han tomado mucha importancia para la formulación de políticas educativas públicas y para generar juicios de calidad educativa a partir de los resultados cuantitativos obtenidos. Estas evaluaciones son de gran valía para los evaluadores y para quienes toman decisiones sobre política pública, quienes valoran los reducidos costos económicos del instrumento, su facilidad de aplicación, la sencillez que ofrecen a la hora de presentar los resultados logrados y su utilidad para la toma de decisiones; poco se debate, en cambio, sobre las limitaciones que pueden presentar estos instrumentos (Herrera Pavo, 2018).

En este sentido, Barrenechea (2010) realiza seis críticas a estas evaluaciones que es necesario tomar en consideración: (1) la tensión que existe entre el concepto de inteligencias múltiples y las evaluaciones estandarizadas, (2) la desatención de componentes del currículo real, (3) los riesgos de enseñar para el test, (4) los incentivos para hacer trampa, (5) la falta de consideración de las diferencias socio económicas de quienes son evaluados y (6) las limitaciones de los resultados de las evaluaciones estandarizadas para predecir el éxito laboral de los estudiantes.

Posiblemente, la más preocupante de estas críticas hechas por Barrenechea (2010) es el riesgo de que las instituciones educativas adapten lo que enseñan a lo que se evalúa en las pruebas estandarizadas, simplificando el currículo oficial y dejando de lado contenidos relevantes para la formación integral de los y las estudiantes. Por ello, "el alineamiento de los estándares [de aprendizaje], los contenidos y la evaluación es crucial" (La Marca, 2001, p. 9), ya que, en ausencia de una fuerte alineación entre el currículo, los estándares empleados para evaluar y las evaluaciones externas, las instituciones educativas pueden ignorar los contenidos del proyecto educativo y enseñar solo lo que se evalúa.

Estudios de alineación

El alineamiento se define como "el grado en el cual los estándares y las evaluaciones están de acuerdo y se sirven conjuntamente para guiar el sistema hacia los estudiantes, para que aprendan lo que se espera que sepan y hagan. Como tal, el alineamiento debe entenderse como la calidad de la relación entre los estándares y la evaluación (Webb, 1999).

En este sentido, a medida que el alineamiento se consolida como método de análisis de las evaluaciones estandarizadas externas, pasa de concebirse como una mera correspondencia entre estándares e ítems de evaluación, a "medir la profundidad y amplitud de los estándares [...] de contenido académico" (Case & Zucker, 2005, p.4), brindando una perspectiva cualitativa sobre el alineamiento de contenidos.

Los métodos más comúnmente empleados para analizar el grado de alineación entre los estándares y el currículo, según Case y Zucker (2005), son tres: (1) desarrollo secuencial, (2) análisis de documentos y (3) 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 que propone la correspondencia secuencial entre estándares y pruebas.

El método basado en análisis de documentos examina, mediante plantillas de especificaciones que codifican su contenido y estructura, las normas y las formas de las pruebas. Las matrices resultantes especifican la correspondencia entre los estándares, los contenidos, los procesos cognitivos y las preguntas de la prueba.

Por último, el método basado en la revisión de expertos analiza el alineamiento entre estándares y evaluaciones, después de que ambos han sido desarrollados.

La integración de estos tres métodos permite a quienes diseñan las políticas públicas, a quienes diseñan las pruebas estandarizadas y a los líderes educativos incrementar la coherencia de los esfuerzos de reforma y rendición de cuentas (Webb, 1997).

A la luz de los resultados de diversos estudios, se han ido generando métodos cada vez más rigurosos para el análisis del alineamiento entre evaluaciones y currículo; según Roach, Niebling y Kurz (2008), se han identificado tres modelos de preferencia como marcos de referencia para el diseño y la implementación de estudios de alineación: (1) El modelo *Surveys of Enacted Curriculum*, desarrollado por Porter y Smithson, que utiliza la ayuda de cuatro expertos para clasificar los estándares y los ítems de acuerdo con criterios de contenido y demanda cognitiva, usando una matriz bidimensional de contenido y un cuestionario de prácticas instructivas en el aula (Porter, 2002). (2) El modelo *Achieve*,

desarrollado en 1998 en el Learning Research and Development Center de la Universidad de Pittsburgh, que emplea cinco criterios para el alineamiento -centralidad del contenido, centralidad del rendimiento, desafío, equilibrio y rango- y se apoya en un grupo de expertos y escalas numéricas para realizar el alineamiento (Case & Zucker, 2005). (3) El modelo Webb, desarrollado por Norman Webb (1997), que analiza con la ayuda de un panel de expertos cinco categorías de alineación -conurrencia de categorías de contenido, profundidad de conocimiento, amplitud de conocimiento, balance de representación y fuente de desafío-.

El modelo Webb es bastante apreciado por las diversas perspectivas de análisis y la sistematicidad de su desarrollo (Gamoran, Porter, Smithson, & White, 1997); ha sido empleado para el análisis del alineamiento, con frecuencia en el área de Matemática, en varios estados de los Estados Unidos de América, incluyendo Puerto Rico; y, en Latinoamérica, además, se ha empleado en Colombia y Honduras (Lopez, 2014; Ramos & Casas, 2016); las constantes revisiones del modelo (N. Webb, Herman, & Webb, 2006; N. L. Webb, 1997, 1999, 2002, 2005, 2009) han permitido su paulatino perfeccionamiento.

El currículo y las evaluaciones estandarizadas en el Ecuador

En el Ecuador, se ha vivido recientemente una transición curricular que nos ha llevado del proyecto de 2010 y de los estándares de aprendizaje de 2012, al currículo y estándares de 2016, ahora vigentes. Este nuevo proyecto cuenta con una estructura y goza de una articulación con los estándares de aprendizaje con los que no cuenta el anterior currículo de 2010, por lo que fueron seleccionados para los fines de este estudio (Herrera Pavo, 2016).

El currículo se describe como un proyecto educativo cuyo fin es "promover el desarrollo y la socialización de las nuevas generaciones" (Ministerio de Educación, 2016, p. 4). En este proyecto, "cada área organiza los aprendizajes [de cada nivel educativo] en bloques curriculares, donde los contenidos de

los aprendizajes son expresados en destrezas con criterios de desempeño” (Ministerio de Educación, 2016, p. 9).

Por su parte, los estándares “son descripciones de los logros de aprendizaje esperados de los estudiantes y constituyen referentes comunes que deben alcanzar a lo largo de su trayectoria escolar” y “están en correspondencia con el currículo oficial” (Ministerio de Educación, 2016, p. 5).

Finalmente, la prueba 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). La evaluación consta de 155 ítems, de los cuales 40 pertenecen al área de Matemática, todos están diseñados con respuesta de opción múltiple, cerrada.

El modelo Webb

Tomando en consideración los elementos de la propuesta curricular vigente en el Ecuador que acabamos de definir, las cinco categorías de análisis del modelo Webb se describen de la siguiente manera:

(1) La concurrencia de categorías de contenido analiza si los ítems de la prueba de evaluación incorporan los mismos contenidos que las destrezas con criterios de desempeño, asociadas a los diferentes estándares de aprendizaje y bloques curriculares. Es decir, revisa si una evaluación representa los mismos contenidos definidos en el currículo.

(2) La profundidad de conocimiento compara la complejidad cognitiva expresada en las destrezas con criterios de desempeño y aquella de los ítems de la prueba de evaluación. Para definir la demanda cognitiva se emplean cuatro niveles de profundidad de conocimiento: memorístico, procesamiento, estratégico y extendido.

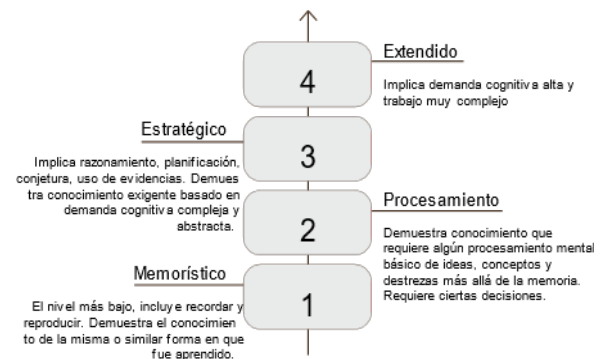


Figura 1: Niveles de profundidad de conocimiento.

(3) La amplitud de conocimiento indica la cobertura que realizan los ítems de la prueba de evaluación de las destrezas con criterios de desempeño del currículo, agrupadas en estándares y bloques curriculares.

(4) El balance de representación indica la distribución de los ítems de la prueba entre las destrezas con criterios de desempeño del currículo, agrupadas en estándares y bloques.

(5) Finalmente, la fuente de desafío determina si los ítems de evaluación de las pruebas estandarizadas tienen algún tipo de problemática conceptual o técnica que dificulte o facilite su resolución por parte de los estudiantes. Hay problema de fuente de desafío si: (a) un estudiante que conoce el contenido del ítem puede escoger una respuesta incorrecta o si (b) un estudiante que no conoce el contenido del ítem puede escoger una respuesta correcta.

Método

Este estudio realizó la validación del modelo de alineación de Webb y sus instrumentos, con técnicas mixtas y con un enfoque descriptivo, a partir de la documentación de los estudios desarrollados en los Estados Unidos de América por Norman Webb (Webb, 1997, 1999, 2002, 2005, 2009).

Para esta validación, se realizó un ejercicio de análisis del alineamiento entre el currículo de Matemática de bachillerato de 2016 y dos formas liberadas de la evaluación Ser Bachiller, régimen Costa, 2017.

En la Tabla 1 se resumen los elementos curriculares analizados: destrezas con criterios de desempeño, por estándar y por bloque curricular.

	Estándares	Número de destrezas con criterios de desempeño
Bloque 1		
Algebra y funciones	E.M.5.1.	8
	E.M.5.2.	11
	E.M.5.3.	36
	E.M.5.4.	9
	E.M.5.5.	14
Bloque 2		
Geometría y medida	E.M.5.6.	17
	E.M.5.7.	6
	E.M.5.8.	4
Bloque 3		
Estadística y probabilidad	E.M.5.10.	15
	E.M.5.11.	4
	E.M.5.9.	6
Total		130

Tabla 1. Desagregación de los componentes curriculares del área de Matemática (bachillerato).

Fases del estudio

Este estudio se desarrolló en diez fases que se describen a continuación.

Fase cero. Selección y contacto con el equipo de expertos.

El primer paso fue seleccionar los expertos del área de Matemática que colaborarían con el estudio. Para ello se realizó una convocatoria abierta y se seleccionó a aquellas personas que cumplieron con los siguientes criterios: (1) poseer un mínimo de 2 años de experiencia docente, (2) ser en la actualidad docente del área de Matemática, coordinador académico o coordinador de área, (3) tener experiencia en la elaboración de pruebas de base estructurada, (4) registrar las mejores calificaciones en las evaluaciones de desempeño de la institución donde labora.

Cuatro expertos que cumplieran el perfil requerido fueron seleccionados y, una vez que aceptaron participar, se les envió a sus correos electrónicos la ficha Información de jueces, en la que se recogieron datos de identificación, formación académica y experiencia profesional.

Fase uno. Entrenamiento de expertos y asignación de responsabilidades.

Se efectuó el entrenamiento individual de expertos para familiarizarlos con la teoría, conceptos claves de alineación, su importancia y las categorías del modelo Webb. A la vez que se les brindó información sobre los compromisos y las tareas que deberían asumir a lo largo de la investigación: (a) Participar en todo el proceso de familiarización y entrenamiento y en las diferentes fases de la investigación. (b) Registrar según su criterio la información solicitada en las fichas de trabajo individual. (c) Colaborar activamente en las sesiones de consenso. (d) Asistir a todas las sesiones de trabajo.

El análisis de alineación desarrollado con los expertos supuso veinticinco horas de trabajo, distribuidas en nueve días. De las nueve sesiones, cinco se basaron en el trabajo individual de análisis y cuatro se dedicaron al logro de consensos. La dinámica propuesta fue la siguiente:

- Las decisiones del panel de expertos se registrarían en fichas denominadas de acuerdo con las categorías de análisis del modelo y un código alfanumérico -#A, #B y #C-, donde el número representaba el orden de la ficha y la letra a los tres bloques curriculares del área de Matemática: Algebra y funciones, Geometría y medida y Estadística y probabilidad. Cada una de las fichas incluiría una descripción de la categoría de análisis, el objetivo y la tarea a cumplir y un espacio destinado al código de experto asignado para el estudio.
- Para los trabajos de consenso se elegiría a un experto como responsable, encargado de registrar las decisiones.
- Adicionalmente, al término de cada encuentro se abriría un grupo de discusión para que los expertos validasen cada instrumento aplicado.

- Las observaciones, sugerencias y comentarios se anotarían en un diario de campo.

Fase dos. Codificación individual de acuerdo con las categorías de contenido.

Cada experto registró la información en tres fichas -2A, 2B y 2C- denominadas Ficha Individual Experto Validador_ Concurrencia en las Categorías de Contenido entre el Currículo de Matemática y Ser Bachiller. El contenido de la ficha incluyó el código y descripción de las destrezas con criterios de desempeño del currículo, un espacio para colocar los códigos de los ítems de la prueba Ser Bachiller que a su juicio correspondían a cada destreza y un espacio para contabilizar los ítems asociados. Para este ejercicio también se entregó un juego de copias de las formas liberadas Ser Bachiller que iban a ser analizadas.

Fase tres. Consenso de acuerdo con las categorías de contenido.

El grupo de expertos discutió la relación de cada ítem de evaluación con las destrezas con criterios de desempeño 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 3 votos, lo que correspondía a la mitad más uno de los expertos del panel.

Las decisiones quedaron registradas en tres fichas -3A, 3B y 3C- denominadas Consenso de Expertos Validadores_ Concurrencia en las Categorías de Contenido entre el Currículo de Matemática y Ser Bachiller, que respondían a una estructura similar a las fichas individuales, contemplando un espacio para registrar el consenso.

Fase cuatro. Codificación individual de acuerdo con la profundidad de conocimiento de las destrezas con criterios de desempeño.

En esta fase los expertos realizaron un entrenamiento de codificación para familiarizarse con los cuatro niveles de profundidad cognitiva descritos por Webb, para ello se codificaron algunos ítems

pertenecientes a formas de años previos liberadas por INEVAL, de tal manera que no se influenciara el criterio de los expertos al momento de realizar el ejercicio definitivo con las formas del régimen Costa de 2017; al finalizar se aclararon dudas.

La información se registró en tres fichas -4A, 4B y 4C- denominadas Individual Experto Validador_ Demanda /Profundidad Cognitiva (Profundidad de conocimiento DOK) _ Currículo. El contenido de la ficha incluyó el código y descripción de las destrezas con criterios de desempeño, un espacio para registrar el nivel cognitivo de dicha destreza y un espacio para anotar el verbo (saber hacer) que le llevó a su decisión, en el caso de ciertas destrezas con más de un saber hacer. Para este ejercicio se facilitó a cada experto un listado con la taxonomía Webb para Matemática.

Pese a que solo se debía asignar un nivel de profundidad de conocimiento a las destrezas con criterios de desempeño identificadas en los consensos de la fase de análisis de concurrencia, se pidió a los expertos codificar las 130 destrezas con criterios de desempeño del currículo de Matemática de bachillerato y se midió el tiempo que les demandó este ejercicio.

Fase cinco. Consenso de acuerdo con profundidad de conocimiento de las destrezas con criterios de desempeño.

Al igual que en el consenso anterior, se discutieron los aportes individuales y se registró el consenso en un máximo de dos rondas, por unanimidad o por mayoría con un mínimo de 3 votos, lo que correspondía a la mitad más uno de los expertos del panel.

En las fichas denominadas Consenso de Expertos Validadores_ Demanda Cognitiva/ profundidad (Profundidad de conocimiento DOK) _Currículo -5A, 5B y 5C-, se registró primero el consenso alcanzado y posteriormente se brindó un espacio para observaciones.

Fase seis. Codificación individual de acuerdo con la profundidad de conocimiento de los ítems Ser Bachiller.

Con la misma dinámica de la fase cuatro, cada experto registró su criterio en las fichas denominadas Individual Experto Validador_ Demanda/Profundidad Cognitiva (Profundidad de conocimiento DOK) _Ser Bachiller - 6A, 6B, 6C-. Estas fichas contenían la numeración y la descripción del ítem de evaluación, un espacio para anotar el nivel de profundidad de conocimiento, otro para asignar la destreza con criterios de desempeño para cada ítem y un último espacio para observaciones.

Fase siete. Consenso de acuerdo con la profundidad de conocimiento de los ítems Ser Bachiller.

Al igual que en los dos consensos anteriores se discutieron los aportes individuales y se registró el consenso en un máximo de dos rondas: por unanimidad o por mayoría con un mínimo de 3 votos, lo que correspondía a la mitad más uno de los expertos del panel. Se emplearon las fichas denominadas Consenso Experto Validador_ Demanda /Profundidad Cognitiva (Profundidad de conocimiento DOK) _Ser Bachiller- 7A, 7B y 7C-, con la misma estructura descrita para las fichas individuales.

Fase ocho. Cobertura de conocimiento con los consensos antes logrados.

En esta fase se solicitó asignar hasta 3 contenidos que vincularan las destrezas con criterios de desempeño y los ítems asociados en las fases previas, con el objetivo de determinar la cobertura y registrar las observaciones en cada caso.

Para ello se empleó la ficha 8, denominada Comparativa de Consenso de Expertos Validadores. Esta ficha contenía el nivel de profundidad cognitiva de las destrezas con criterios de desempeño, seguido del nivel de profundidad cognitiva del ítem asociado a dicha destreza, es decir, en un solo instrumento se ofreció el resumen de los tres consensos anteriores.

El consenso sobre los temas que vinculaban destrezas con criterios de desempeño e ítem de evaluación se realizó con la misma dinámica planteada en los ejercicios anteriores.

El análisis de la fuente de desafío se manejó de manera transversal a lo largo de todas las fases anteriores.

Fase nueve. Entrevista.

En esta fase se entrevistó a los expertos para conocer su interpretación sobre el proceso metodológico empleado para el análisis del alineamiento entre el currículo de Matemática de bachillerato y la evaluación estandarizada Ser Bachiller, así como su percepción de los resultados. Para la entrevista semiestructurada se elaboró una guía a partir de las categorías de análisis del modelo y las diferentes fases de la investigación.

El procesamiento de la información

El procesamiento de la información, que se expone a continuación, se apegó a las directrices dadas por Webb (2005) en el Manual de entrenamiento de la Web alignment Tool (WAT).

(1) Para la concurrencia de las categorías de contenido, se analizó cada una de las decisiones individuales, recogidas en las fichas 2 A, B y C, 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 a 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: (1) Existe alineamiento cuando al menos se codifican, en promedio, seis ítems de evaluación en cada bloque curricular. (2) El alineamiento es débil

si los ítems de evaluación codificados en cada bloque curricular, en promedio, son cinco. (3) No existe alineamiento si los ítems de evaluación codificados en cada bloque curricular, en promedio, son menos de cinco.

Las fichas 3 A, B y C, que recogían los consensos sobre esta categoría, fueron empleadas como base para el desarrollo de la siguiente fase, donde se analizó el nivel de demanda cognitiva de las destrezas con criterios de desempeño identificadas como concurrentes con ítems de las pruebas de evaluación; además, estas fichas de consenso sirvieron para refinar el análisis.

(2) Para la categoría de profundidad de conocimiento o demanda cognitiva, se analizó si un experto, según lo recogido en las fichas 4 A, B y C, y en las fichas 6, A, B y C, 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 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: (1) 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. (2) 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. (3) 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.

Las fichas 5 A, B y C, y las fichas 7, A, B y C, que recogían los consensos de esta categoría, se emplearon para refinar el análisis.

(3) Para la categoría de amplitud de conocimiento, y con el objetivo de establecer en qué medida una prueba cubre 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, recogidas en las fichas 2 A, B y C, a partir de las cuales se calculó el porcentaje, en promedio, del total de destrezas que presentaban 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: (1) 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. (2) 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. (3) 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.

(4) Para la categoría de balance de representación, y con el objetivo de 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, a partir de los registros de las fichas 2 A, B y C,. 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 es:

$$\text{Índice de balance} = 1 - \frac{\left(\sum \left| \frac{1}{(\text{DCD})} - \frac{I_k}{(H)} \right| \right)}{2}$$

$K = 1$

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: (1) Existe alineamiento cuando obtenemos un índice de balance de, al menos, 0.7. (2) El alineamiento es débil si el índice de balance es menor a 0.7, pero igual o mayor que 0.6. (3) No existe alineamiento si el índice de balance está por debajo de 0.6.

(5) Para la categoría de fuente de desafío, que recibió un tratamiento transversal, cada experto registró las observaciones técnicas o conceptuales que fue encontrando en los ítems de evaluación.

Las observaciones técnicas pueden presentarse si: (1) 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. (2) Se presenta algún tipo de información adicional que puede causar algún tipo de confusión. (3) Si la redacción no es adecuada para la población a la que va dirigida la evaluación.

Para el caso de las observaciones de tipo conceptual, estas se presentan si: (1) En el planteamiento del ítem existe algún tipo de error conceptual y en consecuencia no hay respuesta correcta en las opciones o estas son ambiguas. (2) 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.

Estos problemas se discutieron en las sesiones de consenso y fueron registrados en el diario de campo.

Resultados

Atendiendo al objetivo del estudio, se presentan los resultados sobre la experiencia de validación,

relativos al propio modelo, la metodología y los instrumentos empleados.

El modelo Webb

Los expertos estimaron que es factible aplicar el modelo en el Ecuador, ya que proporciona resultados desde varias perspectivas, y destacan su claridad, precisión, orden, sistematicidad y profundidad para el análisis de la relación existente entre los diferentes componentes del currículo y la evaluación.

La capacitación de los expertos

En cuanto al proceso de formación, los expertos consideraron que la sesión de capacitación, aunque muy extensa fue pertinente en cuanto a contenido. Por otro lado, la taxonomía Webb de niveles de profundidad cognitiva para Matemática fue considerada un instrumento práctico por su orden alfabético y por el resumen incluido de los cuatro niveles de profundidad de conocimiento que debían analizar.

La etapa de familiarización fue calificada por los expertos como efectiva para aclarar dudas, acercarse a la teoría, a los conceptos claves de alineación, a la importancia de alinear y profundizar el conocimiento de los criterios del modelo Webb.

La dinámica de trabajo

En cuanto al equipo, el cronograma y la dinámica de trabajo, los expertos consideraron enriquecedora y complementaria la posibilidad de participar de manera individual y grupal. Por otro lado, estimaron que el número de sesiones individuales y de consenso, 5 y 4 respectivamente, permitieron cumplir el cronograma de trabajo de manera muy ajustada.

En este sentido, el tiempo asignado para cada tarea fue cronometrado para realizar futuros ajustes en el cronograma. El tiempo empleado para el consenso de concurrencia de categorías de contenido fue mayor en el análisis de los primeros ítems y se redujo en los últimos.

Un número par de expertos ayudó a promover el debate en las sesiones de consenso, ante los

frecuentes casos de tablas en las codificaciones individuales del grupo; aunque, en ocasiones, llegaba a prolongar estas sesiones en exceso. No obstante, la designación de un responsable para registrar la información en estas sesiones facilitó la dinámica de trabajo.

En cuanto a la comunicación con el equipo de expertos, el correo electrónico y la mensajería instantánea fueron medios eficaces.

Los instrumentos de registro

Respecto a la validación de los instrumentos y medios de recolección de la información en la aplicación del modelo Webb, los expertos, independientemente de la estrategia empleada para registrar la información, cumplieron con la tarea de la jornada debido a la claridad de las indicaciones. Es decir, las fichas cumplieron con el criterio de objetividad.

La extensión de las denominaciones de las fichas de recolección de la información en las distintas fases de trabajo impidió que los expertos se familiarizaran rápidamente con ellas.

La ficha de información de jueces causó confusión en dos preguntas referentes a la experiencia y aspectos profesionales, debido a su planteamiento ambiguo y poco preciso.

En cuanto a la ficha de concurrencia de categorías de contenido los expertos recomiendan: resaltar visualmente los contenidos conceptuales en las destrezas con criterios de desempeño, además de añadir un espacio para observaciones. Se recomienda, además, incluir un área para registrar el código de la destreza con criterios de desempeño en las formas de evaluación analizadas.

Las fichas para las sesiones de consenso de categorías de contenido y nivel de profundidad de conocimiento requieren modificar su estructura para que se visibilice en ellas, además del resultado del consenso, las decisiones individuales de cada experto.

La estructura de los ítems Ser Bachiller al ser trasladados a la ficha individual para registrar

el nivel de profundidad cognitiva resulta poco practica por la cantidad de hojas que involucra.

La ficha comparativa de consensos favoreció para que los expertos tengan una visión global del trabajo hasta ese momento realizado; en su estructura hay necesidad de incluir a modo de encabezado, los nombres de cada columna para facilitar el trabajo.

La información de fuente de desafío requiere ser registrada en una ficha y sesión específica de trabajo.

La entrevista fue muy provechosa en cuanto a la información que proporcionó; la guía de entrevista facilitó un dialogo fluido entre el entrevistador y el entrevistado, pero limitó la interacción entre los expertos.

El análisis de los datos

Todos los datos fueron procesados mediante hojas de cálculo, y este procedimiento nos permitió generar matrices de procesamiento de datos que pueden ser empleadas en próximas investigaciones.

Aunque no es el objetivo de este estudio, se ofrecen los resultados del procesamiento de los datos (Tabla 2), que muestran que el alineamiento de currículo y la evaluación externa para el área de Matemática en el nivel de bachillerato es aceptable en cuanto a demanda cognitiva, pero presenta debilidad en algún bloque en cuanto a la concurrencia de categorías de contenidos y balance y es inaceptable en cuanto a la amplitud de rango. No entraremos, sin embargo, en la interpretación de estos resultados.

En cualquier caso, los coeficientes de correlación intraclase calculados para las codificaciones de la categoría de profundidad de conocimiento solo son aceptables para las destrezas con criterios de desempeño (0,742), no así para los ítems de evaluación (0,468).

Bloque	Concurrencia	Profundidad Cognitiva	Amplitud / Rango	Balance
Algebra y Funciones	Aceptable	Aceptable	Inaceptable	Débil
Geometría y medida	Débil	Aceptable	Inaceptable	Aceptable
Estadística y probabilidad	Aceptable	Aceptable	Inaceptable	Aceptable

Tabla 2: Resumen de los resultados de alineación para el área de Matemática.

Discusión

Nuestro trabajo ha tratado de resolver la falta de información con respecto a la metodología y los instrumentos empleados en un estudio de alineamiento basado en el modelo Webb en el contexto Latinoamericano. Hemos sido meticulosos en el diseño y validación de la metodología y los instrumentos, ya que de su calidad dependerá que se obtengan resultados fiables en investigaciones sobre alineamiento posteriores que usen este modelo.

Del trabajo realizado con los expertos en este estudio se extrae la utilidad del modelo para analizar el alineamiento en el contexto ecuatoriano. Además, este análisis se ve facilitado por la estructura del currículo 2016 y su articulación con los estándares de aprendizaje (Herrera Pavo, 2016), un ejercicio similar con el currículo de 2010 hubiera supuesto un significativo trabajo previo de organización de las destrezas con criterios de desempeño en bloques disciplinares, estableciendo su relación con los estándares de aprendizaje.

En cuanto a la metodología, es necesario destacar la importancia de la selección del equipo de expertos con los criterios apropiados. Así como tomar en consideración las limitaciones para formar un panel de expertos docentes del nivel bachillerato y personal experto en evaluación, si no se cuenta con apoyo institucional, por ejemplo, del Ministerio de Educación y del propio Instituto Nacional de Evaluación.

Además, la capacitación es clave, ya que, tal y como advierte La Marca, con respecto al

proceso de codificación, “por su naturaleza subjetiva, este tipo de revisión requiere de un fuerte componente de capacitación” (2001, p. 3). Los expertos han ratificado que los distintos momentos de capacitación los ayudaron a familiarizarse con la metodología propuesta, aumentaron su conocimiento, posibilitaron la unificación de criterios y fundamentaron sus decisiones.

El coeficiente de correlación intraclase nos ayuda a determinar la fiabilidad de las decisiones tomadas por cada experto durante el ejercicio de codificación; en nuestro caso, los resultados fueron aceptables para la codificación de las destrezas con criterios de desempeño, pero no para los ítems de evaluación asociados. Pensamos que para mejorar la fiabilidad del estudio es preciso contar con un número mayor de expertos en el panel, y reforzar el proceso de capacitación.

Por otro lado, es necesario considerar la necesidad de mantener una comunicación muy cercana con los expertos, para lo que el correo electrónico y la mensajería instantánea son herramientas clave.

Las sesiones de trabajo permitieron cumplir el cronograma propuesto, ya que, al revisar el modelo y el volumen de los documentos a analizar, currículo y evaluaciones, se ponderaron los tiempos y se ajustó el cronograma. Sin embargo, considerando sesiones de trabajo de jornada completa se pueden reducir las actuales 9 sesiones, y además se aprovecharía de mejor manera el tiempo de los expertos, proporcionándoles la oportunidad de intercambiar ideas en los respectivos debates con mayor profundidad.

En este sentido, es preciso destacar la relevancia de las fases de consenso, ya que “tomar en consideración los desacuerdos de los expertos cambia las conclusiones sobre el alineamiento, no solo mostrando un alineamiento más débil del previamente demostrado, sino también cambiando los perfiles de alineamiento sobre, por ejemplo, la cobertura relativa de estándares

específicos” (N. Webb, Herman, & Webb, 2006, p. 1).

De este modo, aunque se hacen los cálculos de alineamiento en función de la codificación individual de los expertos para cada categoría, es preciso revisarlos desde una perspectiva cualitativa basada en las discusiones y resultados de la fase de consensos.

En cuanto a los instrumentos, los expertos destacaron su validez, sugiriendo modificaciones para facilitar la tarea de codificación.

Finalmente, pensamos que la fase de entrevistas, que se llevó a cabo al final del estudio, contribuyó a clarificar el proceso de investigación e identificar situaciones que pueden obstaculizar o mejorar su desarrollo. Además, profundizar en la perspectiva de los codificadores, ayuda a comprender mejor los datos obtenidos. Por ello, recomendamos reconocer la importancia de esta fase, que se podría complementar con grupos de discusión para favorecer el contraste de puntos de vista.

Conclusiones

Es importante considerar que este es el primer intento de validación de instrumentos de alineación según el modelo Webb realizado en el Ecuador, y que ha sido posible gracias a la estructura del currículo 2016, alineado con los estándares de aprendizaje.

El desarrollo de este estudio ha permitido tener una perspectiva clara sobre cómo implementar un análisis del alineamiento entre el currículo de Matemática de bachillerato y la evaluación estandarizada Ser Bachiller, siguiendo del modelo Webb.

Este estudio ha demostrado la validez de la metodología y los instrumentos del modelo Webb para el análisis del alineamiento del currículo, los estándares de aprendizaje y las pruebas de evaluación en el contexto ecuatoriano.

La aplicación del modelo Webb es factible, debido a que presenta de manera sistemática la metodología, existen herramientas electrónicas

y bibliografía de consulta para la formación de los expertos. Además, se trata de un modelo que combina técnicas cualitativas y cuantitativas, que favorecen que se pueda corroborar la fiabilidad de los datos recabados mediante los instrumentos.

Aunque los expertos concluyeron que todos los instrumentos propuestos para las sesiones de trabajo individual favorecen el proceso de codificación de acuerdo con lo establecido en cada categoría de alineación y sugirieron algunas mejoras que serán implementadas, es importante tener en cuenta que para que el análisis se desarrolle de la manera más adecuada es preciso un minucioso proceso de selección de los expertos, un número mayor de expertos en el panel y un proceso de capacitación riguroso.

Por último, es preciso considerar que los datos cuantitativos proporcionados por la aplicación del modelo deben ser revisados a la luz de la información de tipo cualitativo obtenida de las discusiones que tienen lugar en las sesiones de consenso, ya que esta información será valiosa para hacer precisiones sobre los resultados del estudio de alineamiento.

Referencias

- Barrenechea, I. (2010). Evaluaciones estandarizadas: seis reflexiones críticas. *Education Policy Analysis Archives/ Archivos Analíticos de Políticas Educativas*, 18(8), 1–27. Retrieved from <http://www.redalyc.org/html/2750/275019712008/>
- Case, B., & Zucker, S. (2005). Methodologies for Alignment of Standards and Assessments. In *China-US Conference on Alignment of Assessments and Instruction* (p. 5). Retrieved from https://images.pearsonassessments.com/images/tmrs/tmrs_rg/AlignmentMethodologies.pdf?WT.mc_id=TMRS_Methodologies_for_Alignment
- Gamoran, A., Porter, A. C., Smithson, J., & White, P. A. (1997). *Upgrading High School Mathematics Instruction: Improving Learning Opportunities for Low-Achieving, Low-Income Youth. Educational Evaluation and Policy Analysis*, 19(4), 325–338. <https://doi.org/10.3102/01623737019004325>
- Herrera Pavo, M. Á. (2016). Hacia un nuevo currículum abierto y flexible, articulado con los estándares de aprendizaje. In S. Endara Rosales (Ed.), *Educación, calidad y buen vivir* (pp. 191–200). Azogues: UNAE. Retrieved from https://www.academia.edu/29721691/Hacia_un_nuevo_currículum_abierto_y_flexible_articulado_con_los_estándares_de_aprendizaje
- Herrera Pavo, M. Á. (2018). La evaluación estandarizada de aprendizajes en Ecuador sus limitaciones y la necesidad de plantear alternativas. In *Primer encuentro de metodologías para la evaluación de la educación* (pp. 49–60).
- INEVAL. (2017). *Ser Bachiller 2017. Ficha técnica y conceptual*. Ineval. Quito. Retrieved from http://www.evaluacion.gob.ec/evaluaciones/wp-content/uploads/2017/07/Ineval_fichaSBAC17_20170224.pdf
- La Marca, P. M. (2001). Alignment of Standards And Assessments as an Accountability Criterion. *Practical Assessment, Research & Evaluation*, 7(21), 1/4. Retrieved from <https://pareonline.net/getvn.asp?v=7&n=21>
- Lopez, A. A. (2014). 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*, 19(2). <https://doi.org/10.7203/relieve.19.2.3024>
- Martinic, S. (2010). La evaluación y la reforma educativa en América Latina. *Revista Iberoamericana de Evaluación Educativa*, 3(3).
- Ministerio de Educación. (2016). *Currículo de los niveles de educación obligatoria*. Quito: Ministerio de Educación. Retrieved

- from <https://educacion.gob.ec/wp-content/uploads/downloads/2016/08/Curriculov2.pdf>
- Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3–14. <https://doi.org/10.3102/0013189X031007003>
- Ramos, L., & Casas, L. (2016). Demanda cognitiva en estándares educativos y evaluación en álgebra. In A. Jiménez, J. L. González, M. T. Sánchez, P. Hernández, C. Fernández, F. J. Ruiz, ... A. Berciano (Eds.), *Investigación en Educación Matemática XX* (p. 633). Málaga: SEIEM. Retrieved from <http://funes.uniandes.edu.co/8961/1/Casas2016Demanda.pdf>
- Roach, A. T., Niebling, B. C., & Kurz, A. (2008). Evaluating the alignment among curriculum, instruction, and assessments: Implications and applications for research and practice. *Psychology in the Schools*, 45(2), 158–176. <https://doi.org/10.1002/pits.20282>
- Webb, N., Herman, J., & Webb, N. L. (2006). *Alignment of Mathematics State-level Standards and Assessments: The Role of Reviewer Agreement*. Los Ángeles. Retrieved from <https://cresst.org/wp-content/uploads/R685.pdf>
- Webb, N. L. (1997). Determining Alignment of Expectations and Assessments in Mathematics and Science Education. *NISE Brief*, 1(2), 1–8. Retrieved from http://archive.wceruw.org/nise/Publications/Briefs/Vol_1_No_2/NISE_Brief_Vol_1_No_2.pdf
- Webb, N. L. (1999, August). *Alignment of Science and Mathematics Standards and Assessments in Four States*. Research Monograph No. 18. Wisconsin: Wisconsin Center for Education Research, 1025 W. Johnson Street, Madison, WI 53706 (\$7.75). Tel: 608-265-9698. Retrieved from <https://eric.ed.gov/?id=ED440852>
- Webb, N. L. (2002). An Analysis of the Alignment Between Mathematics Standards and Assessments for Three States. In American Educational Research Association Annual Meeting in New Orleans, Louisiana April 1-5, 2002 (p. 17). Wisconsin. Retrieved from http://facstaff.wceruw.org/normw/AERA_2002/Alignment_Analysis_three_states_Math_Final_31502.pdf
- Webb, N. L. (2005). *Web Alignment Tool. Training Manual*. Wisconsin.
- Webb, N. L. (2009). *Design of Content Alignment Studies in Mathematics and Reading for 12th Grade NAEP and other Assessments to be used in Preparedness Research Studies*. Retrieved from <https://www.nagb.gov/content/nagb/assets/documents/publications/design-document-final.pdf>
- Cómo citar: (Norma APA)
- Figuerola Chávez, J. F., & Herrera Pavo, M. Ángel. (2019). 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*, 4(4). <https://doi.org/10.33386/593dp.2019.4.91>