

Universidad de la República
Comisión Coordinadora del Interior

**EVALUACIÓN DIAGNÓSTICA DEL NIVEL DE APRENDIZAJE EN
LECTURA Y MATEMÁTICA AL MOMENTO DE INGRESO A LA
UNIVERSIDAD DE LA REPÚBLICA**

Marco de referencia para la evaluación de Matemática

Abril 2017

Emiliano Clavijo, Tabaré Fernández, Verónica Figueroa, Agustina Marques,
Cecilia Rodríguez y Pilar Rodríguez

El objetivo de la Evaluación Diagnóstica en el área de Matemática es describir el nivel de conocimientos matemáticos con los que cuentan los estudiantes que ingresan a la Universidad de la República. *“El conocimiento matemático es esencial para la preparación de los jóvenes para la vida en la sociedad moderna, ya que una creciente proporción de problemas y situaciones que se presentan en el trabajo y contextos profesionales, requieren razonamiento de cierto nivel, utilización de estrategias y herramientas matemáticas”* (ANEP 2011). En este sentido, se entiende que la competencia matemática es determinante para las trayectorias de los estudiantes por la Educación Superior, independientemente de la carga matemática de la carrera elegida, entendiendo que la aplicación de estas herramientas es clave para resolver todo tipo de problemas y para el desempeño personal, académico y profesional.

Partimos de la noción más amplia y contemporánea dada a la competencia matemática en ocasión de la revisión del marco conceptual de PISA en 2012, que fue nuevamente empleada en PISA 2015. Según éste, aquella es *“la capacidad de un individuo para formular, emplear e interpretar las matemáticas en una variedad de contextos. Incluye razonar matemáticamente y el uso de conceptos matemáticos, procedimientos, datos y herramientas para describir, explicar y predecir fenómenos. Asiste a los individuos a reconocer el rol de las matemáticas en el mundo y a hacer valoraciones y a tomar decisiones bien fundadas, necesarias para los ciudadanos constructivos, comprometidos y reflexivos.”* (OCDE 2016).

Este documento busca plasmar las decisiones metodológicas tomadas al realizar la Evaluación Diagnóstica en el área de Matemática.

Las evaluaciones en Matemática

A nivel internacional, puede decirse que existen dos grandes enfoques teóricos sobre la evaluación en Matemática que desarrollan pruebas con amplia cobertura. Ellos son el Programa para la Evaluación Internacional de Estudiantes (PISA) y Tendencias en el Estudio Internacional de Matemática y Ciencia (TIMSS)¹.

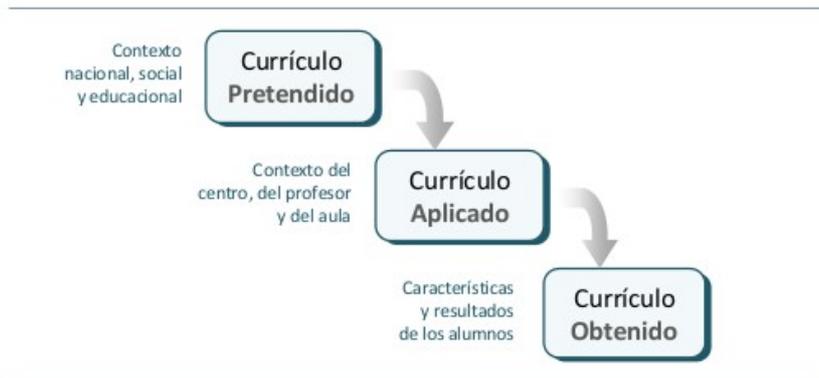
TIMSS² utiliza el **currículo como principal concepto organizador** de sus pruebas, entendiendo que este modelo contempla el contexto de los alumnos en dos niveles: el nacional, social y educacional, así como el contexto del centro educativo, el profesor y el aula. En este sentido, las evaluaciones de TIMMS vienen acompañadas de exhaustivos informes sobre los currículos pretendidos en cada país y de información cuantitativa y cualitativa brindada por profesores en cuanto a su formación y experiencia, así como sobre enfoques aplicados en la enseñanza, actitudes desde la docencia, recursos disponibles en las aulas y los centros, etc. A su vez, incluye el indicador de resultados, el currículo aprendido por los estudiantes. El modelo se presenta en la siguiente figura. El énfasis y el fundamento de toda la evaluación en el currículum se sostiene en la hipótesis en que el principal factor que diferencia el aprendizaje entre los estudiantes, una vez que se controlan los

1 En 2015, como antes en 2003, coincidieron ambos programas internacionales. TIMSS evaluó estudiantes de 4º y 8º grado en 49 países en tanto que PISA lo hizo en una muestra de 540 mil estudiantes de 15 años de edad residentes en 72 países.

2 <http://www.iea.nl/timss>

antecedentes sociofamiliares, es el currículum a que está expuesto en su escuela. De acuerdo a la bibliografía especializada, el currículum incide de manera decisiva sobre lo que aprenden los alumnos (Valverde et al. 2002).

Figura I. El modelo curricular de TIMMS



Fuente: TIMMS 2012

Las evaluaciones orientadas al currículum aprendido por los alumnos, se conforman con consideración en dos dimensiones: por un lado el contenido, que refiere a los diferentes dominios de la matemática, y por el otro, la dimensión cognitiva que especifica los comportamientos esperados de parte de los estudiantes (TIMSS 2012). Las dimensiones evaluadas se presentan en la tabla siguiente.

Tabla I. Marco de evaluación TIMSS

Contenidos	Procesos cognitivos
Números (naturales, fracciones, proporción, modelos y relaciones)	Conocer (recordar, reconocer, calcular, medir, clasificar)
Formas y mediciones geométricas (puntos, líneas, ángulos, formas bidimensionales y tridimensionales)	Aplicar (seleccionar, representar, poner en práctica, resolver problemas rutinarios)
Representación de datos (lectura e interpretación, organización y representación)	Razonar (analizar, generalizar/especializar, integrar/sintetizar, justificar, resolver problemas no rutinarios)

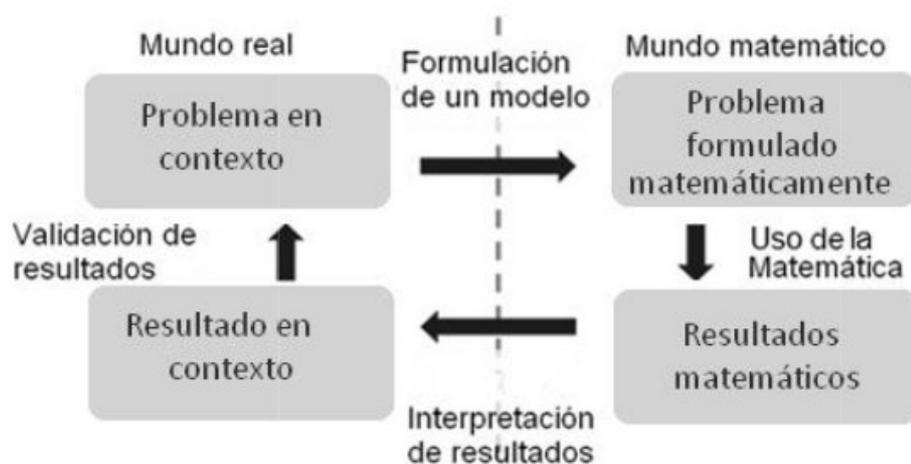
Fuente: Elaboración propia en base a TIMSS (2016)

Cabe destacar que las pruebas TIMSS se realizan en dos niveles: para estudiantes en su cuarto año de Educación Primaria y luego en su octavo año de educación, lo que en Uruguay corresponde a segundo año de Ciclo Básico. Naturalmente, esto representa dos pruebas con sus diferencias adaptadas a cada nivel. En este caso, nos concentraremos en la prueba aplicada a los estudiantes del 8º grado, que se corresponde en Uruguay al 2º año de Educación Media Básica. La prueba tiene un total de 14 cuadernillos, en el que cada cuadernillo consta de 14 ítems de Matemática y 14 de Ciencias, respetando la distribución de categorías de contenido y de procesos cognitivos del pool entero de ítems. Esta amplitud de posibilidades viene dada por la aplicación sistemática de la Teoría de Respuesta al Ítem, estándar mundial en materia de evaluación de aprendizajes.

En cuanto a las Pruebas PISA³, su evaluación en Matemática se centra en torno a la idea de **competencia**. Este concepto se encuentra presente en distintos marcos curriculares formulados desde fines de los años noventa. Desde esta época, tal enfoque se ha ido extendiendo primero en la Unión Europea, donde los países han pasado a establecer un currículum alejado de un listado de contenidos sino formulado en competencias básicas que los estudiantes deberían alcanzar al finalizar el ciclo de estudios secundarios. Entre ellas se encuentra la competencia matemática. Tal es definida como *la habilidad para utilizar y relacionar los números, sus operaciones básicas, los símbolos y las formas de expresión y razonamiento matemático, tanto para producir e interpretar distintos tipos de información, como para ampliar el conocimiento sobre aspectos cuantitativos y espaciales de la realidad, y para resolver problemas relacionados con la vida cotidiana y con el mundo laboral*. También, supone aplicar destrezas y actitudes que permiten razonar matemáticamente, comprender una argumentación matemática, expresarse y comunicarse en el lenguaje matemático (ANEP 2011).

Las nociones de *modelo* y de ciclo de *construcción de modelos* son esenciales al concepto de PISA sobre la competencia matemática, y tienen como supuesto un individuo que se plantea y resuelve en forma activa los problemas que enfrenta. El siguiente esquema resume esta noción.

Figura II. El modelo de competencia matemática en PISA



Fuente: ANEP (2011)

3 <https://www.oecd.org/pisa/pisa-2015-results-in-focus-ESP.pdf>

En este sentido, vemos que la noción de competencia matemática difiere a la propuesta por TIMSS en dos aspectos. Por un lado, el concepto organizador de las pruebas PISA es la competencia y no el contenido curricular en los currículos prescriptos por cada país. La segunda diferencia está dada por su enfoque al contexto social y educativo del alumno. TIMSS hace énfasis en los marcos nacionales y de los centros educativos al contemplar el currículo en sus tres niveles (prescripto, desarrollado y aprendido). En cambio, PISA se concentra mayormente en los resultados (las competencias adquiridas por el alumno), entendiendo que hay que situar estas habilidades en contextos tanto en la escuela como fuera de la escuela, pues el uso de la competencia en diferentes contextos sería el valor intrínseco fundamental de la Matemática. Por ello, los ítems de PISA no contemplan solo las dimensiones de contenido y procesos cognitivos, sino que agregan una tercera dimensión a considerar de contextos y situaciones, como indica la siguiente tabla.

Tabla II. Marco de evaluación PISA

Contenidos	Procesos cognitivos	Contextos y situaciones
Cantidad	Reproducción (operaciones matemáticas simples)	Personal
Espacio y Forma	Conexión (relacionar ideas para resolver problemas)	Educativo y laboral
Cambio y Relaciones	Reflexión (razonamiento matemático en sentido amplio)	Público
Incertidumbre		Científico

Fuente: Elaboración propia en base a OCDE (2016)

La validez de todo instrumento de medición se evalúa siempre según su objetivo. En este sentido, no es posible comparar estrictamente las pruebas TIMSS con las pruebas PISA, en tanto una enfoca la evaluación en el currículo y la otra en las competencias adquiridas por la vida adulta. PISA evalúa más allá de los contenidos curriculares específicos, prestando atención sobre todo a los conocimientos y destrezas relevantes para la adultez, lo que también incluye ciertas capacidades transversales, como comunicación, flexibilidad, uso de TICs, etc. (Acevedo 2005). Este foco de PISA es lo que le permite concentrarse en competencias más “universales” que sean comparables en diferentes contextos, regiones y países, mientras que TIMSS sí se enfoca en los acontecimientos dentro del centro educativo, con una impronta institucional y local. También es necesario comprender que PISA pretende más la evaluación del rendimiento de los sistemas educativos respecto a sus objetivos subyacentes –tal como los define la sociedad– que con referencia en la enseñanza y el aprendizaje de un cuerpo disciplinar de conocimientos (Acevedo 2005).

De esta forma, es importante señalar que ambas pruebas distan de ser contrapuestas, sino que se complementan. Ambos enfoques son importantes para evaluar el conocimiento matemático de los alumnos, entendiendo en primer lugar, que la evaluación en el contexto del centro educativo condiciona las competencias que adquieren para la vida adulta, y en segundo lugar, que si bien las competencias matemáticas son las que definen el uso efectivo de la matemática como herramienta para la vida personal, académica y laboral, éstas no pueden adquirirse desprovistas de contenido.

Marco adoptado en la Evaluación Diagnóstica de la CCI

En el apartado anterior se contrastaron a modo de orientación, los modelos de dos pruebas externas internacionales que evalúan conocimientos matemáticos en adolescentes. En este presentaremos los principales rasgos del proceso de elaboración del marco conceptual que orientó la evaluación diagnóstica en Matemática.

El primer elemento a tener en cuenta es la CCI adoptó y es una extensión del proyecto que se venía implementando en el Centro Universitario de la Región Este (CURE) desde el año 2011 (Rodríguez et al. 2012). Uno de sus fundamentos conceptuales se ubica en el College Board de Estados Unidos, entidad que ha diseñado y pautado hace décadas orientaciones cruciales en la enseñanza matemática de las escuelas de aquel país (Madison et al. 2003).

El segundo elemento relativo a la formulación del marco tiene que ver con el procedimiento seguido. Desde su comienzo, la conformación del banco de ítems estuvo orientada por el objetivo de cubrir los contenidos curriculares de los programas de las distintas asignaturas “Matemática” que se dictan en los Bachilleratos (Diversificados y Tecnológicos) del país. *“Se partió de la revisión de los programas curriculares de Matemática de los últimos dos años de bachillerato, estableciéndose como referencia de competencia mínima los contenidos de los programas de las orientaciones de bachillerato con menos contenidos de Matemática”*(Rodríguez 2017). Se conformó un grupo de discusión en el que participaron cinco docentes de la Educación Media Superior. La intersección de contenidos entre estos programas fue utilizada como las competencias y contenidos estándar que debería poseer cualquier estudiante que haya acreditado la Educación Media Superior. Es decir, se fijaron como estándares de contenidos, aquellos que podrían ser esperables de un estudiante que hubiera transitado el Bachillerato SocioHumanístico o el Bachillerato Tecnológico en Turismo, que históricamente ha tenido la menor carga de Matemática. Es de notar incluso que hasta la reforma curricular de 2006 de la Enseñanza Secundaria, la orientación humanística carecía de tales cursos en el último grado para la orientación “Derecho y Ciencias Sociales”. Este sería el equivalente al “currículo pretendido” propuesto por TIMMS.

En tercer lugar, la operacionalización de estas definiciones usó como antecedente la prueba Scholastic Assessment Test (SAT). Esta es una prueba con más de 80 años de trayectoria dirigida a estudiantes de bachillerato con el fin de evaluar su preparación académica y es la más comúnmente utilizada por las universidades estadounidenses. Cabe destacar que estas pruebas no cumplen una función diagnóstica de los estudiantes al ingreso a la Educación Superior, sino que determinan las posibilidades académicas de los estudiantes que las toman, por lo que los bachilleres en Estados Unidos se preparan intensamente para dar estas pruebas. Otro aspecto relevante es que los ítems no se presentan contextualizados, esto es insertos en una situación problemática que se propone emular un problema “extra-escolar” o típico de la “vida diaria privada o laboral”. También es una característica de esta formulación que no existen ítems encadenados o derivados de un mismo texto: a diferencia de PISA, la actividad es el ítem y este podría decirse que es estrictamente “escolar”.

El cuarto elemento es que nuevamente este umbral estuvo presente en la evaluación de los ítems generados, esto es, una vez que estaban seleccionados y antes de ser incorporados al banco de ítems, pasaron por un proceso de “jueces” que calificaron su dificultad frente a los estándares definidos.

De esta forma, la ED en el área de Matemática permitiría describir a los estudiantes que ingresan frente a estándares basados en el currículum de los Bachilleratos con menor contenido y exigencia Matemática, situación que debe ser tenida presente al analizar los resultados, en particular si

se tiene en vista aquellos para una carrera de alta exigencia matemática. Estas distancias con las exigencias de las ingenierías o con las ciencias económicas es un correlato que debe ser tenido presente.

El marco de referencia de Matemática se organiza en torno a cuatro dimensiones que contemplan los SAT para evaluar la dimensión matemática: i) Números y Operaciones; ii) Álgebra; iii) Geometría; y iv) Análisis de datos, Estadística y Probabilidad.

Es importante destacar que si bien hubo un proceso de definición de estándares de contenido, no están clasificados por procesos cognitivos tal como se presenta en PISA o TIMSS, por lo que la elección de los ítems y la conformación de la prueba responde a una sola dimensión, la de los contenidos (aún si se consideran los niveles de dificultad atribuidos).

Elaboración de los ítems de Matemática

Uno de las etapas más importantes que debe seguir un proyecto de evaluación de aprendizajes es la construcción de un banco de ítems, teóricamente fundados en el marco de referencia propuesto, técnicamente formulados y validados por especialistas, dado que, como se mencionara, la calidad de los ítems de una prueba depende fundamentalmente de su validez de contenido (Pérez Juste, 2006). Por este motivo y considerando la falta de recursos específicos disponibles para realizar tan amplia tarea, se decidió apelar a ítems ya testeados en otras evaluaciones, como por ejemplo de las pruebas SAT desarrollada por The College Board en Estados Unidos.

En su inicio, la UAE del CURE conformó un primer listado con 192 ítems que luego de una primera y preliminar revisión quedó reducido a 186. Estos fueron enviados a cuatro evaluadores que valoraron la representación exhaustiva de los contenidos, su redacción, su adecuación en relación con los estándares de contenido previamente establecidos y los distractores, indicando en una planilla su puntuación en una escala del 1 al 4 donde 1 es “nada pertinente”, 2 “poco pertinente”, 3 “algo pertinente” y 4 “muy pertinente”. Además se solicitó que se realizaran comentarios si fuera necesario como por ejemplo, respecto a su dificultad, el lugar óptimo de su ubicación en una secuencia de prueba escrita, etc.

Al recibirse las puntuaciones y comentarios de cada evaluador se procedió de la siguiente forma: los comentarios se copiaron para tenerlos en cuenta en caso que el ítem pasara la evaluación; las puntuaciones fueron ingresadas en una planilla sumándose los puntajes dados a cada ítem por los cinco evaluadores. Se consideraron como ítems aptos los que obtuvieron el rango de puntajes más altos. Si había una gran dispersión entre las valoraciones se apeló a los comentarios para decidir si el ítem se consideraba apto.

A su vez, respecto de los ítems utilizados en 2015, se calcularon los parámetros de discriminación y dificultad con base en la Teoría de Respuesta al Ítem, a modo de poder filtrar ítems con indicadores psicométricos débiles para la conformación del banco de ítems 2016, al igual que para la distribución de los mismos entre los cuadernillos.

En 2016 el proceso de evaluación de nuevos ítems fue coordinado por el Departamento de Matemática y Aplicaciones del CURE y participaron como evaluadores docentes con experiencia previa en la redacción de ítems. Se seleccionaron 91 nuevos ítems y se enviaron a los evaluadores el

documento con los estándares de contenido, las pautas de evaluación y los ítems seleccionados. Este grupo evaluador también sugirió una nueva secuencia de los ítems en los tres cuadernillos.

A su vez, el procedimiento de “ajuste” del banco de ítems con base en indicadores psicométricos se repitió en 2017 para el 2016, resultando el banco final de ítems a febrero de 2017 en un conjunto de 90 ítems de matemática validados tanto conceptual como psicométricamente. Estos ítems se pueden clasificar por dimensión evaluada, lo que se presenta en la siguiente tabla.

Tabla III. Distribución de los ítems del banco de Matemática según dimensión evaluada

Dimensiones	Número total	Porcentaje
Números y Operaciones	23	26%
Álgebra	40	44%
Geometría	18	20%
Análisis de datos, estadística y probabilidad	9	10%
Total	90	100%

Fuente: Elaboración propia según el banco de ítems 2016

Elaboración de la prueba de Matemática

Para cada una de las cuatro dimensiones se definió cuál sería su proporción en el total de la prueba conforme a la preponderancia que se les da en los programas curriculares a esos contenidos. Se determinó que cada categoría podía tener una proporción de ítems distribuidos según la siguiente tabla:

Tabla IV. Dimensiones de contenidos matemáticos y proporción de ítems en la prueba

Dimensiones	Proporción de ítems
Números y Operaciones	15-18%
Álgebra	45-52%
Geometría	15-20%
Análisis de datos, estadística y probabilidad	9-14%

Fuente: Elaboración propia

El objetivo de tener varios cuadernillos responde a la necesidad de maximizar la cobertura de los ítems del banco por un lado, sin perder de vista que suficientes estudiantes deben responder a un ítem para asegurar la fiabilidad de su medición. La adjudicación de los cuadernillos a los estudiantes fue aleatoria, permitiendo suponer que son poblaciones comparables. Cada cuadernillo posee 44 ítems, extensión que se mantuvo para todas las ediciones de la evaluación. En las evaluaciones de 2014 y 2015 se utilizaron dos cuadernillos. Para 2016 se agregó un tercero con ítems seleccionados por los evaluadores. Este nuevo cuadernillo sigue la proporción de ítems para cada categoría descrita en la Tabla V.

En línea con la bibliografía especializada (Kolen & Brennan, 2014) y al igual que en Lectura, dentro del banco de ítems se definió un subconjunto compuesto por aquellos que fueron aplicados en 2014 y que con base en sólidos índices psicométricos, podían funcionar como “anclas” en 2015, haciendo posible la equiparación y el escalamiento de los dos cuadernillos de pruebas confeccionados. Estos 17 ítems ancla volvieron a utilizarse con el mismo fin en 2016 para los tres cuadernillos de ese año. Se distribuyen en las diferentes categorías evaluadas en línea con la proporción estipulada para cada una, como indica la tabla a continuación.

Tabla V. Categoría de los ítems ancla en la prueba de Matemática

Categoría	Número de casos
Números y Operaciones	3 (18%)
Álgebra	8 (47%)
Geometría	4 (23%)
Análisis de datos, estadística y probabilidad	2 (12%)
Total	17 (100%)

Fuente: Elaboración propia según el banco de ítems 2016

Referencias bibliográficas

Acevedo Díaz, José Antonio (2005) "TIMSS Y PISA. DOS PROYECTOS INTERNACIONALES DE EVALUACIÓN DEL APRENDIZAJE ESCOLAR EN CIENCIAS" *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, Vol. 2, No 3, pp. 282-301

Administración Nacional de Educación Pública (2011)

Kolen, Michael J. & Brennan, Robert L. (2004) "Test Equating, Scaling, and Linking: Methods and Practices" *Springer Science & Business Media*

Madison, B., Briggs, A., Choike, J., Halvorsen, K., Kennedy, Daniel, Walters, T. (2003) "The College Board Mathematical Sciences Framework" *Boston, MA. Obtenido de http://www.collegeboard.com/prod_downloads/about/association/academic/math3.pdf*

Mullis, I. V. S., Cotter, K. E., Fishbein, B. G., & Centurino, V. A. S. (2016) "Developing the TIMSS 2015 achievement items" en **M. O. Martin, I. V. S. Mullis, & M. Hooper** (Eds.) "Methods and Procedures in TIMSS 2015" *Boston College, TIMSS & PIRLS International Study Center*

OECD-PISA (2013) "PISA 2012 Assessment and Analytical Framework. Mathematics, Reading, Science, Problem Solving and Financial Literacy" *Paris: OECD Publishing. Obtenido de <http://dx.doi.org/10.1787/9789264190511-en>*

OECD (2016), PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematics and Financial Literacy, PISA, OECD Publishing, Paris

Pérez Juste, Ramón (2006) "EVALUACION DE PROGRAMAS EDUCATIVOS", *Editorial LA MURALLA*

Programme for International Student Assessment 2012

Rodríguez, P. (2017) "Creación, Desarrollo y Resultados de la Aplicación de Pruebas de Evaluación basadas en Estándares para Diagnosticar Competencias en Matemática y Lectura al ingreso a la Universidad" *Revista Iberoamericana de Evaluación Educativa* , 89-107.

Rodríguez, P., Correa, A., & Díaz, A. (2012) "Informe sobre los resultados de la Evaluación Diagnóstica 2012" *Maldonado: Universidad de la República-CURE. Obtenido de www.cure.edu.uy/index.php/uae*

Trends in International Mathematics and Science Study (2012) "TIMSS 2011 Marcos de la evaluación", *Instituto Nacional de Evaluación Educativa, Ministerio de Educación, Cultura y Deporte de España en cooperación con TIMSS & PIRLS International Study Center, Boston College*

Trends in International Mathematics and Science Study (2016) "TIMSS 2015 Assessment Frameworks", *International Association for the Evaluation of Educational Achievement (IEA), TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College*

Valverde, G. A., Bianchi, L. J., Wolfe, R. G., Schmidt, W. H., & Houang, R. T. (2002) "According to the Book - Using TIMSS to investigate the translation of policy into practice through the world of textbooks" *Dordrecht: Kluwer*