



CUARTO INFORME DEL ESTADO DE LA EDUCACIÓN

Informe Final

Costa Rica en las pruebas PISA 2009 de Competencia Lectora y Alfabetización Matemática

Investigadoras:

Eiliana Montero

Shirley Rojas

Evelyn Zamora

Ana María Rodino

2012



CONSEJO NACIONAL DE RECTORES

Nota: Las cifras de las ponencias pueden no coincidir con las consignadas por el IV Informe Estado de la Educación en el tema respectivo, debido a revisiones posteriores. En caso de encontrarse diferencia entre ambas fuentes, prevalecen las publicadas en el Informe.

Contenido

Resumen Ejecutivo	3
Hallazgos relevantes	4
Recomendaciones.....	7
Introducción.....	8
Pruebas estandarizadas y PISA: Una breve inducción.....	12
Características de la prueba PISA que la hacen una evaluación única en su género	15
Los informes de resultados generales de PISA 2009.....	17
Algunas limitaciones de PISA	18
Marco general de referencia: algunos determinantes del rendimiento académico	20
Factores de contexto del país.....	21
Características del estudiante	22
Características del centro educativo	27
Características del docente.....	35
Estudios realizados en Costa Rica	38
Representatividad de la muestra de PISA en Costa Rica	40
Análisis correlacionales y de regresión multinivel con los resultados PISA 2009 en Alfabetización Matemática y Competencia Lectora: variables predictoras ...	43
Análisis bivariados	44
Modelo de regresión multinivel para identificar variables predictoras de los puntajes de PISA en Competencia Lectora y Alfabetización Matemática	49
Justificación.....	49
Metodología	52
Resultados.....	66
Recomendaciones para acciones de política educativa	74
Bibliografía.....	77

Resumen Ejecutivo

Las pruebas PISA (Programme for International Student Assessment) son exámenes estandarizados comparativos internacionales. Abarcan tres áreas: Alfabetización Matemática, Competencia Lectora y Alfabetización Científica, pero no miden contenidos curriculares específicos sino más bien la habilidad para usar el conocimiento en la resolución de ítems que presentan situaciones o problemas de la vida diarias.

En 2009 se realizaron en Costa Rica las pruebas PISA en dos áreas: Competencia Lectora y Alfabetización Matemática, en las que los resultados reportados fueron relativamente bajos. La presente investigación busca presentar y discutir resultados clave obtenidos por nuestro país en dichas pruebas, incluyendo factores explicativos asociados al rendimiento de los estudiantes en estos exámenes. A partir de estos hallazgos se discuten además sus importantes implicaciones para acciones concretas de política educativa.

Descriptores

Pruebas PISA, Competencia Lectora, Alfabetización Matemática, colegios públicos, colegios privados, Modelo de Regresión Multinivel, SPSS, STATA.

Hallazgos relevantes

1- Sólo un 53% de los jóvenes de 15 años del país están siendo representados por la muestra de PISA del año 2009, las exclusiones las componen tres grupos, a saber: el sorprendentemente alto porcentaje de estudiantes que poseen adecuaciones en los colegios, aquellos que están en educación abierta y los que no están activos en el sistema educativo. Este es uno de los porcentajes de representación más bajos de los países participantes de PISA en ese año. Para establecer una referencia comparativa, los porcentajes de representación para México y Chile son respectivamente 61% y 85%.

Es de esperar que el grupo de jóvenes excluidos del marco muestral tenga un desempeño relativamente menor en estas pruebas, comparado con el grupo que sí se muestreó.

En una simulación realizada para establecer cuáles podrían ser los resultados de Costa Rica en las dos pruebas bajo estudio, Competencia Lectora y Alfabetización Matemática, si nuestro país tuviera, respectivamente, los niveles de representación de México y de Chile, los datos arrojaron que, en el caso de la comparación con México, se mantendrían la diferencias significativas a favor de Costa Rica que se presentan cuando se comparan los datos observados en la prueba de Competencia Lectora, mientras que también se mantendría la diferencia estadísticamente significativa, favorable a México, en Alfabetización Matemática.

Por el contrario en la comparación con Chile, si Costa Rica exhibiera los mismos niveles de representación en PISA que este país, los promedios de rendimiento en ambas pruebas serían significativamente más bajos para Costa Rica, o sea los resultados de Costa Rica quedarían por debajo de los de Chile.

2- Los desempeños generales de la muestra costarricense de PISA 2009 en las dos pruebas estudiadas, Competencia Lectora y Alfabetización Matemática, son relativamente bajos.

En Competencia Lectora se definen 7 niveles de desempeño, siendo los más bajos 1b y 1a, y 6 el más elevado.

Para esta prueba la categoría que agrupa el mayor porcentaje de estudiantes costarricenses (35%) es el nivel 2, mientras que el porcentaje acumulado, es decir los estudiantes que están en ese nivel o en niveles más bajos es de 67%.

Los desempeños típicos del nivel 2 en Competencia Lectora se describen a continuación:

Algunas tareas de este nivel requieren que el lector localice una o más piezas de información, que tal vez deban deducirse y es posible que tengan que cumplir varias condiciones. Otras requieren reconocer la idea principal del texto, comprender relaciones o analizar el significado dentro de una parte limitada del texto cuando la información no es prominente y el lector debe realizar deducciones de bajo nivel. Las

tareas de este nivel pueden comprender comparaciones o contrastes basados en una única característica del texto. Las típicas tareas de reflexión de este nivel requieren que los lectores realicen comparaciones o varias conexiones entre el texto y conocimientos externos, recurriendo a experiencias y actitudes personales.

En contraste, el nivel 6, el más alto de la escala, no fue alcanzado por ninguno de los estudiantes de la muestra costarricense y se describe a continuación:

Las tareas de este nivel normalmente requieren que el lector realice múltiples deducciones, comparaciones y contrastes, detallados y precisos. Requieren la demostración de una comprensión plena y detallada de uno o más textos y pueden implicar la integración de información de más de un texto. Es posible que las tareas requieran que el lector aborde ideas que le son poco familiares, en presencia de una información prominente en conflicto, y que genere categorías abstractas para su interpretación. Las tareas de *reflexión y evaluación* pueden requerir que el lector plantee hipótesis o evalúe críticamente un texto complejo sobre un tema poco conocido, teniendo en cuenta múltiples criterios o perspectivas y aplicando una sofisticada comprensión más allá del texto. Una condición destacada de las tareas de *acceso y recuperación* de este nivel es la precisión del análisis y la atención minuciosa a detalles poco aparentes en los textos.

Ahora bien, en el caso de Alfabetización Matemática los niveles también son 7, definidos desde 0 (por debajo del nivel 1) hasta 6.

La moda (valor más frecuente) en este constructo para la muestra costarricense es el nivel 1, con 33% de los estudiantes, mientras que los ubicados en el nivel 0 son 24%. Así, más de un 50% de los examinados se encuentran en las dos categorías más bajas. Estas se describen a continuación:

Nivel 0: Los estudiantes en este nivel no tienen éxito en tareas matemáticas más básicas medidas por PISA. Su patrón de respuestas hace prever que no serían capaces de resolver la mitad de las tareas en una prueba elaborada exclusivamente con preguntas del nivel 1. Estos estudiantes tienen serias dificultades para utilizar la matemáticas en cualquier contexto.

Nivel 1: Los estudiantes de este nivel pueden responder a las preguntas que involucren contextos familiares, donde toda la información relevante está presente y las preguntas están claramente definidas. Estos estudiantes son capaces de identificar la información para llevar a cabo los procedimientos rutinarios de acuerdo a las instrucciones directas en situaciones explícitas. Pueden realizar acciones obvias y continuar inmediatamente con determinados estímulos.

Por otra parte, el nivel 6, el más alto de la escala, que ninguno de los estudiantes de la muestra costarricense logró alcanzar, se define como sigue:

En este nivel el estudiante puede contextualizar, generalizar y utilizar la información basándose en sus investigaciones o en la modelización de una situación o problema complejo. El estudiante puede relacionar la información proveniente de diferentes

fuentes o representaciones. El estudiante debe ser capaz de realizar razonamientos matemáticos avanzados, puede aplicar los conocimientos, la simbología y la matemática formal para entender y explicar nuevas situaciones y generar nuevas estrategias para dar respuestas a las mismas. En este nivel el estudiante puede expresarse en forma precisa, reflexiona sobre sus hallazgos e interpreta y reflexiona apropiadamente.

3- Uno de los puntos que más llama la atención en las comparaciones realizadas según categorías de población, se refiere a las diferencias, que sin temor a exagerar podemos llamar enormes, entre los promedios de estudiantes de colegios públicos y privados. La diferencia, favorable a los de colegios privados, es 87,72 en Competencia Lectora Lectura y 79,63 en Alfabetización Matemática. Estas magnitudes son superiores, en ambos casos, a la desviación estándar de la distribución total de los puntajes. De manera que estas disparidades tan dramáticas parecen entregar una nueva señal sobre lo que puede llamarse la “crisis de la educación secundaria pública” y sobre las desigualdades, cada vez mayores, en oportunidades educativas que se evidencian en nuestro país.

4- Sin embargo, es importante enfatizar que las diferencias encontradas entre estudiantes de colegios públicos y privados no se deben, per se, a un efecto intrínseco del colegio, sino a un cúmulo de variables que se relacionan con ventajas y desventajas educativas dentro y fuera del aula, y que están también asociadas con el tipo de colegio al que asiste el estudiante. Estos hallazgos fueron posibles gracias al uso de la regresión multinivel. Algunas de las variables identificadas son más susceptibles a la posibilidad de intervención educativa que otras.

5- De hecho, en los modelos multinivel que se estimaron para identificar factores asociados al rendimiento en las pruebas, los indicadores de dimensiones de lectura (un constructo que sí puede ser objeto de intervención con acciones de política educativa) presentan evidencia sólida de un importante poder predictivo para el desempeño en ambos exámenes.

6- En concordancia con lo anterior, las siguientes recomendaciones específicas se derivan de los resultados de las regresiones multinivel. Las variables correspondientes mostraron asociaciones de relevancia tanto con el puntaje en Competencia Lectora como con el puntaje en Alfabetización Matemática.

Enseñar explícitamente en el aula las estrategias que son eficaces para resumir un texto.

Enseñar explícitamente en el aula las estrategias que son eficaces para comprender un texto.

Diseñar estrategias pedagógicas que promuevan en los estudiantes una actitud positiva hacia la lectura.

Fomentar la lectura “en línea”.

Impulsar la lectura por iniciativa propia.

Crear en el centro educativo condiciones que generen menos dificultades para desarrollar la enseñanza, remediando aspectos como la falta de profesores calificados y la escasez o inadecuación del material pedagógico.

7- Con el objeto de enfatizar la importancia de tomar acciones concretas de política educativa para promover las dimensiones de lectura, se realizó una simulación con el modelo multinivel para establecer en cuánto disminuirían las diferencias entre estudiantes de colegios públicos y privados si los de colegios públicos mostraran los mismos valores promedio en los indicadores de lectura que muestran los estudiantes de colegios privados.

Los resultados fueron contundentes, pues si ése fuera el caso la diferencia monumental que se observa entre estos dos grupos se reduciría dramáticamente. La reducción sería más de un 67% de la diferencia promedio observada actualmente en el caso de Competencia Lectora y más de un 70% de la diferencia de promedios observada en el caso de Alfabetización Matemática.

Entonces la brecha desmedida que se evidencia hoy entre estas dos categorías de estudiantes podría decrecer de manera muy notable si se atacara la problemática asociada a los bajos niveles en los indicadores de lectura que presentan, en promedio, los estudiantes de colegios públicos.

Nótese además, que estos indicadores fueron relevantes no solo para predecir los puntajes en la prueba de Competencia Lectora, sino también para la de Alfabetización Matemática, lo que sugiere que las acciones programáticas deberían ser transversales para las diversas materias del currículo.

Recomendaciones

1-Es esencial que el país continúe participando en PISA y contar así con datos longitudinales que permitan dar seguimiento a la evolución del rendimiento de nuestros estudiantes en el tiempo. Esto posibilitará a su vez el monitoreo de los posibles impactos de acciones programáticas y pedagógicas con propósitos de mejoramiento.

2-Las bases de datos generadas por PISA brindan un mundo de posibilidades diversas de análisis para los investigadores en educación y evaluación educativa, pero se requiere de equipos de investigación que cuenten no solo con marcos conceptuales apropiados, sino también con conocimientos relativamente avanzados en modelos multivariados de análisis de datos aplicados a estudios observacionales en las Ciencias Sociales, como regresión multinivel y modelos de ecuaciones estructurales. Es importante que el Estado de la Educación fomente la conformación de estos equipos interdisciplinarios para la realización de estos estudios.

3-Contando con los datos de PISA 2012, es imperativo realizar un nuevo estudio con regresión multinivel longitudinal, incorporando como segunda medición los puntajes en PISA 2012, para generar una primera estimación de la evolución o cambio en el período y de los factores que explican este cambio.

4-De igual forma, se debe continuar explorando las posibilidades de los Modelos de Ecuaciones Estructurales o (SEM por sus siglas en inglés), para buscar evidencias empíricas de planteamientos causales complejos derivados de marcos conceptuales actualizados sobre determinantes del rendimiento en estas evaluaciones y considerando también los factores propios del contexto local.

5- Al realizar comparaciones con los desempeños de otros países o regiones se debe considerar el sesgo que se genera cuando los porcentajes de representación de PISA son muy disímiles. En tal caso las comparaciones directas entre puntajes observados no son válidas, y se debería utilizar la metodología descrita en la sección de este capítulo titulada “Representatividad de la muestra de PISA en Costa Rica”.

Introducción

El objetivo fundamental de este capítulo especial es presentar y discutir resultados clave obtenidos por nuestro país en las pruebas PISA 2009 de Lectura y Matemática, incluyendo factores explicativos asociados al rendimiento de los estudiantes en estos exámenes. A partir de estos hallazgos se discuten además sus importantes implicaciones para acciones concretas de política educativa.

De las tres áreas presentes en PISA, Competencia Lectora, Alfabetización Matemática y Alfabetización Científica se decidió estudiar y analizar los resultados de Costa Rica en las dos primeras. La razón principal para incluir Competencia Lectora es que este constructo fue precisamente el énfasis de la aplicación de PISA en el 2009, tal como se explica más abajo en este capítulo central y con mucha mayor amplitud en el documento titulado “La competencia lectora de los estudiantes costarricenses según la evaluación internacional PISA 2009+”, de Ana María Rodino, que también acompaña a este capítulo. En el caso de Alfabetización Matemática se consideró que dada la actual coyuntura educativa, con una reciente reforma de los programas de estudio de Matemática, orientados, al igual que PISA, por el enfoque de resolución de problemas, era muy pertinente una reflexión sustantiva sobre los resultados, la que se presenta en otro de los documentos que componen este capítulo cuyo autor es Oscar Salas y titulado “Constructo ‘Alfabetización Matemática’, según PISA”.

Además de analizar y discutir desde cada disciplina específica, Lingüística y Enseñanza de la Matemática los resultados descriptivos de PISA para nuestro país, la identificación de variables clave que explican o predicen los resultados, es uno de los objetivos esenciales de este capítulo, porque precisamente de este análisis de factores asociados al rendimiento en PISA, utilizando los modelos de regresión multinivel es

posible brindar recomendaciones sustentadas de manera sólida a nivel científico para acciones de política educativa que permitan mejorar los resultados.

Asimismo, como dimensión complementaria a la anterior, se presentarán también resultados provenientes de las pruebas de diagnóstico de III Ciclo en Español y Matemática elaboradas por el Ministerio de Educación Pública (MEP), así como variables del estudiante, docente y centro educativo que predicen el desempeño en dichas pruebas diagnósticas.

Se debe enfatizar que este documento incorpora además contribuciones producto del intercambio con diversas audiencias interesadas e interlocutores, que han permitido enriquecer estos análisis e interpretaciones con sus sugerencias, recomendaciones y críticas.

Las pruebas PISA (Programme for International Student Assessment) son exámenes estandarizados comparativos internacionales, construidos y administrados por un consorcio que originalmente estuvo formado por los países de la OECD, pero que paulatinamente se amplió, y ya para el año 2009 abarcó 65 países o economías. Su aplicación inicia en el año 2000. Nuestro país participa por primera vez en 2009 y nuevamente en 2012. A nivel latinoamericano los otros países que participaron en PISA 2009 fueron: México y Chile (ambos miembros de la OECD), Brasil, Uruguay, Perú, Panamá, Colombia y el estado de Miranda en Venezuela (Walker, 2011).

Las pruebas PISA abarcan tres áreas: Alfabetización Matemática, Competencia Lectora y Alfabetización Científica, pero no miden contenidos curriculares específicos sino más bien la habilidad para usar el conocimiento en la resolución de ítems que presentan situaciones o problemas de la vida diarias. Cada aplicación de PISA se concentra en un área de contenido, brindando siempre información básica, pero más general, para las otras dos. En el año 2009 el énfasis estuvo en competencia lectora. (Walker, 2011). El énfasis permite brindar información más detallada sobre los componentes específicos del constructo bajo medición, lo cual se logra introduciendo más ítems en la prueba, para generar así una medición más precisa de estos componentes.

Según sus propias palabras, PISA evalúa la aplicación de conocimientos en competencia lectora, alfabetización matemática y alfabetización científica a la resolución de problemas, dentro de un contexto de vida real y situaciones cotidianas (OECD, 2009). Utiliza el término “alfabetización” para denotar su amplio foco sobre la aplicación de conocimientos y destrezas, en contraste con otras pruebas internacionales como TIMSS, que presentan un vínculo más fuerte con marcos de referencia curriculares y buscan medir el dominio de los estudiantes de conocimientos, destrezas y conceptos específicos. (Walker, 2011). El contenido de las pruebas PISA es sacado de amplias áreas de contenido, por ejemplo, espacio y forma para Matemática, en contraste con contenidos más específicamente basados en el currículo, como geometría o álgebra. Así, PISA evalúa en qué grado los estudiantes, cerca del final de su educación obligatoria, han adquirido el conocimiento y destrezas para participar de manera completa y exitosa en la sociedad moderna (Walker, 2011; OECD, 2009).

Se trata, fundamentalmente, de pruebas de papel y lápiz cuya población meta son los estudiantes de 15 años de edad. Más precisamente, la edad de los examinados debe ubicarse entre los 15 años y 3 meses y los 16 años y dos meses. Cada examinado toma un examen de dos horas que incluye los tres componentes, Competencia Lectora, Alfabetización Matemática y Alfabetización Científica. La evaluación incluye una combinación de preguntas de escogencia única y de respuesta abierta en cada una de las tres áreas. Para la calificación de los ítems de respuesta abierta se requiere la colaboración de calificadoros con alto nivel de capacitación y un diseño que permita controlar el efecto del calificador. Aunque a partir del año 2009 se cuenta con un componente de lectura digital, Costa Rica no tomó este componente en esa oportunidad.

Cada ítem en PISA está circunscrito en una de cuatro situaciones: personal, educativa/laboral, pública y científica. (OECD, 2009).

Los puntajes en las escalas de PISA representan niveles a lo largo de un continuum de destrezas de alfabetización. PISA brinda rangos para esos niveles de desempeño, asociados a los puntajes, que describen lo que un estudiante típicamente puede hacer en cada nivel. (OECD, 2009). Tanto Lectura como Matemática en PISA 2009 establecieron 6 categorías para esos niveles de desempeño, con la salvedad de que en Competencia Lectora el nivel 1 se divide en dos partes, a y b. Las interpretaciones son entonces referidas a criterios. En este tipo de modelos los puntajes de la prueba se interpretan en términos de estándares absolutos de desempeño, en contraposición al modelo referido a normas, en donde el significado de la puntuación se basa en la posición relativa del examinado, comparado con el resto de los que tomaron la prueba en su grupo de referencia (Gronlund & Linn, 1990; Montero, 2008). La interpretación de criterios es entonces la más apropiada cuando se trata de pruebas educativas con fines diagnósticos, como en el caso de los exámenes de PISA.

Los estudiantes que rinden la prueba completan, además, un cuestionario que recoge información sobre diferentes variables que podrían estar asociadas con su rendimiento en el examen. El director(a) del centro educativo también llena un cuestionario sobre diversas variables que dan cuenta de diversas características del contexto de su institución y de su propia gestión como director. En general, la administración de estos cuestionarios tarda de 20 a 30 minutos.

Concatenando los datos de los cuestionarios de estudiantes y directores con los de los resultados de las pruebas es posible generar estudios de gran valor, que aportan al conocimiento de los factores del estudiante y de la institución que predicen el rendimiento en los exámenes. Además, la identificación de variables asociadas al puntaje en las pruebas permite generar recomendaciones específicas de política educativa con el propósito de mejorar los niveles de desempeño.

En nuestro país, y siguiendo los protocolos generales de PISA, se tomó una muestra aleatoria estratificada de conglomerados que estuvo constituida 181 instituciones, de ellas 156 eran públicas y 25 eran privadas o subvencionadas, 149 académicas y 32 técnicas, 118 urbanas y 63 rurales, 175 diurnas y 6 nocturnas.

En cada uno de los colegios seleccionados se eligió al azar, partir del listado general de todos los jóvenes de 15 años, una muestra de 30 estudiantes, independientemente del nivel en que estuvieran matriculados.

La aplicación operativa de PISA 2009 se realizó en Costa Rica en el primer semestre del 2010, formando parte de un grupo de 10 países que administraron tardíamente los exámenes y a los que se denomina PISA 2009+.

En total se administraron a la muestra bajo estudio 13 cuadernillos diferentes de examen. Con un complejo diseño de equating (equiparación) se logra controlar el efecto de las diferencias de dificultad en las formas de examen. (OECD, 2012). En este conjunto de 13 cuadernillos hubo 7 que son de menor dificultad, comparados con los administrados en los países de mayor desarrollo y que se ofrecieron como opción a los países que habían logrado un puntaje promedio en la escala de lectura de 450 o menos, o a los países nuevos (como Costa Rica) que, de acuerdo con sus respuestas a la aplicación piloto del 2008, se esperaba que obtuvieran valores en ese rango. La aplicación de estos cuadernillos de menor dificultad se realiza con el propósito de mejorar la precisión de las estimaciones en los niveles de desempeño relativamente más bajos de la escala. (OECD (2012). Por otra parte los métodos de equiparación de puntajes permiten eliminar el efecto de las diferencias de dificultades en las estimaciones del nivel de desempeño.

Para minimizar los efectos de posibles sesgos culturales o de contexto, todos los ítems que se incluyeron en estos cuadernillos fueron previamente revisados y valorados por personal técnico del Departamento de Evaluación Académica y Certificación de la Dirección de Gestión y Evaluación de la Calidad del Ministerio de Educación. Se realizaron los cambios o modificaciones necesarias para reemplazar aquellos reactivos donde se sospechó alguna posibilidad de sesgo por cultura, contexto o vocabulario (P. Mena, comunicación personal, junio 2012).

Las inferencias de la muestra a la población de interés (todos los estudiantes de 15 años del sistema educativo) solo son estadísticamente válidas si se realizan a nivel global, o sea, para todo el país, de manera que no es posible generar inferencias válidas a nivel desagregado, por colegio o por región educativa.

Un corolario de lo anterior es que en países en vías de desarrollo, como Costa Rica, donde un porcentaje relevante de la población en este grupo de edad no está dentro del sistema educativo, los resultados de PISA presentan un sesgo de sobre-estimación, pues los jóvenes que no participan del proceso educativo, o que aun estando en el sistema educativo, forman parte, por otras razones, del marco muestra de PISA, presentarán, probablemente, valores más bajos en los puntajes de estas pruebas, comparados con los muchachos que sí están dentro del marco muestral del cual se obtiene la muestra. Por tanto, es relevante indicar que los resultados obtenidos por Costa Rica no son representativos de toda la población de 15 años cubierta por el sistema educativo y tampoco representan los niveles de competencia en estas áreas o constructos de todos los jóvenes de 15 años. Este punto se ilustra y discute con mayor

profundidad en la sección titulada “Representatividad de la muestra de PISA en Costa Rica”.

En este marco de referencia, consideramos que la verdadera riqueza de PISA y su valor agregado no está en el “ranking” ni en las comparaciones específicas con países que tienen puntajes más altos o más bajos que el nuestro, sino en poder evaluarnos sustantivamente a nosotros mismos con un instrumento de alta calidad, que permite monitorear nuestro desempeño a través del tiempo y que tiene gran relevancia educativa.

Para finalizar esta sección debemos enfatizar que estos análisis de ninguna manera son exhaustivos ni agotan todas las posibilidades de estudio y discusión que ofrecen los datos de la aplicación de PISA en nuestro país. Por el contrario, se trata de un primer intento de acercamiento, que es parcial e incompleto por naturaleza. Por esto esperamos que este primer esfuerzo constituya también una invitación a todos los interesados en la problemática educativa de nuestro país a que realicen sus propias contribuciones.

Pruebas estandarizadas y PISA: Una breve inducción

Una prueba estandarizada es un instrumento de medición que sigue rigurosos procesos de construcción y validación. En general, son exámenes de administración masiva que se aplican y se califican bajo estrictos protocolos que contribuyen a garantizar la comparabilidad de las puntuaciones y la validez de las inferencias acerca del desempeño de los examinados.

Dado que precisamente los temas de interés en educación abarcan la definición y medición de constructos complejos tales como rendimiento académico, habilidad intelectual, conocimientos, actitudes, valores y competencias, es necesario utilizar los conceptos y herramientas de la psicometría y diversos modelos analíticos derivados, que permitan generar indicadores empíricos del grado de validez y confiabilidad de las mediciones realizadas con los instrumentos estandarizados (Martínez et al, 2006; Nunnally & Bernstein, 1995).

Entonces ya sea que la prueba sea de alto impacto para el examinado (como certificación de su aprendizaje) o de bajo impacto (como diagnóstico) las consecuencias de su aplicación y su potencial efecto sobre la vida de las personas o sobre decisiones clave de política educativa, hacen necesario que se deba contar con la participación de profesionales en las áreas de medición educativa y psicometría para que, con su conocimiento especializado, dirijan su proceso de creación y validación y contribuyan a recolectar información empírica y evidencia científica sólida acerca de la confiabilidad y validez de las mismas.

Existe una larga tradición científica en la psicometría y la medición educativa que nos brinda un cuerpo de teoría y métodos para poder construir y validar este tipo de instrumentos. Entre otras cosas, estos procesos científicos de recolección de evidencias

incluyen aplicaciones piloto con el objeto de validar los ítems o reactivos y obtener evidencias de su calidad técnica por medio de modelos como la Teoría Clásica de los Tests (TCT) o más recientemente modelos de Rasch. De acuerdo con los principios que dictan estas disciplinas, y también por razones éticas, solo cuando se cuente con estas evidencias podrían ser estas pruebas usadas para la toma de decisiones.

El tema de la equidad es fundamental y para ello estas disciplinas también cuentan con herramientas técnicas para evitar los sesgos por origen o “background” de los examinados y para que las pequeñas diferencias de dificultad en diferentes versiones de la prueba sean eliminadas y no afecten la comparabilidad de las puntuaciones (procedimientos de “equating” o equiparación de puntajes).

Además de todos estos aspectos técnicos y científicos está el tema de la seguridad y confidencialidad de estos exámenes y los bancos de reactivos o ítems que los componen, y también la transparencia e imparcialidad en su manejo, esto es parte de la experiencia que deben aportar los profesionales en medición y evaluación educativa encargados de dirigir su proceso de construcción y validación.

Las evidencias de la calidad de una prueba estandarizada deben documentarse empíricamente de manera objetiva y clara, y estar científicamente fundamentadas. En este sentido se diferencian de las pruebas construidas por los docentes en el aula, puesto que estas últimas responden a un contexto específico de aprendizaje y a una mediación pedagógica y dinámicas de clase particulares que el docente considera al construir sus exámenes y, que en el caso de las estandarizadas, no pueden controlarse ni considerarse por ser administradas en contextos educativos diversos. En el cuadro 1 se presentan algunas diferencias clave entre las pruebas estandarizadas y las pruebas construidas en el aula.

Cuadro 1
Diferencias entre pruebas educativas estandarizadas y pruebas construidas en el aula ¹

Aspecto	Estandarizadas	Aula
Resultados aprendizajes contenidos medidos	Resultados de medición y contenidos deben ser comunes para todos los centros educativos. Las pruebas rara vez enfatizan aspectos locales del currículo	Consideran aspectos específicos del aprendizaje generado con un cierto proceso de mediación pedagógica y contexto particular de dinámica del aula
Calidad de los ítems	En general, si se siguen los procedimientos establecidos por la psicometría y la medición educativa, la calidad técnica es alta. Los ítems son escritos por especialistas, se realizan pruebas piloto y se seleccionan en términos de su efectividad para realizar una medición adecuada	La calidad de los ítems es desconocida, o al menos no suele estar documentada. Típicamente la calidad es más baja que la de las estandarizadas, por el poco tiempo que dispone el docente y la falta de conocimientos especializados
Confiabilidad	Alta confiabilidad, comúnmente es mayor a 0.8 (tanto en la Teoría Clásica de los Tests como para el modelo de Rasch)	Usualmente desconocida, puede ser alta si el test es cuidadosamente construido, rara vez se calculan medidas empíricas

Aspecto	Estandarizadas	Aula
Administración calificación	Procedimientos estandarizados (uniformes). Se brindan instrucciones específicas. Existen protocolos de administración y calificación	Se favorecen los procedimientos uniformes, pero hay laxitud. Muchas veces no hay protocolos o son poco específicos
Interpretación puntajes	Los puntajes pueden ser comparados con grupos normativos o con niveles de desempeño (enfoque criterial). Existen manuales de la prueba y otras guías que ayudan en la interpretación y el uso	Comparaciones de los puntajes e interpretaciones se limitan a la situación local del centro educativo o el aula
Número examinados	Aplicación masiva, usualmente varios cientos o miles	Aplicación reducida, grupos pequeños en un contexto circunscrito

1/ Adaptado de Gronlund N.E. & Linn R.L. (1990). *Measurement and Evaluation in Teaching*. New York: Macmillan.

El proceso de recolección de evidencias empíricas para la validación de un instrumento estandarizado implica normalmente la consulta a jueces expertos, pero usualmente esto no es suficiente para generar evidencia de validez sólida y suficientemente creíble, hace falta al menos una aplicación piloto y un análisis psicométrico básico del instrumento y de los ítems que lo componen. Entre los métodos y modelos de análisis que se utilizan en este proceso se pueden mencionar los siguientes:

- Análisis de factores exploratorio y confirmatorio
- Teoría Clásica de los Tests (TCT)
- Teoría de Respuesta a los Items (TRI)
- Modelo de Rasch
- Teoría G (Generalizabilidad)
- DIF (Funcionamiento Diferencial del Item)
- Equiparación de puntajes

El análisis de factores, tanto exploratorio como confirmatorio, se refiere a técnicas multivariadas que nos permiten explorar la dimensionalidad subyacente en los datos, supuesto fundamental para la gran mayoría de los modelos de medición que aplicamos para generar evidencias de la confiabilidad y la validez de la prueba (Martínez et al, 2006; Nunnally & Bernstein, 1995).

Por su parte, la Teoría Clásica de los Tests (TCT) es el más antiguo y conocido modelo de medición, que permite generar indicadores numéricos de la calidad técnica del instrumento, incluyendo el coeficiente Alfa de Cronbach, su resultado de mayor importancia, indicador que mide la precisión de la prueba en cuanto a consistencia interna y apunta hacia el grado de estabilidad de los puntajes (Muñiz, 2003).

Con los modelos TRI (Teoría de Respuesta a los Items) y Rasch se obtienen estimaciones de los parámetros del ítem que son menos dependientes de la muestra de examinados y estimaciones de los niveles del constructo en los evaluados que son menos dependientes de la muestra particular de ítems aplicada. Además, en estos modelos existe una estimación específica del error de medición para cada puntaje en la prueba (a diferencia de la TCT donde se asume que el error es constante) (Martínez et al, 2006; Montero, 2001).

En el caso del modelo de Rasch, las estimaciones de las habilidades de los examinados y la dificultad de los ítems están en las mismas unidades de medición (propiedad de medición conjunta), característica que resulta sumamente atractiva a nivel aplicado y de interpretación sustantiva, pues permite evaluar el desempeño del examinado con un enfoque referido a criterios, es decir, valorando en términos absolutos lo que puede o no hacer, en contraposición a las interpretaciones de normas que tradicionalmente eran las únicas permitidas con el modelo de medición más antiguo, la Teoría Clásica de los Tests (Bond & Fox, 2001; Prieto & Delgado, 2003).

PISA usa precisamente el modelo de Rasch para generar y validar los niveles de desempeño que define para la interpretación cualitativa que se realiza a partir de los puntajes numéricos obtenidos en las pruebas. Los programas TIMSS, SERCE y TERCE también utilizan el modelo de Rasch, gracias a esta misma propiedad de medición conjunta. Para conocer un poco más las características de este modelo los lectores pueden referirse al recuadro titulado “El modelo de Rasch: una herramienta esencial para la evaluación educativa de gran escala”, de Shirley Rojas.

Los temas de Funcionamiento Diferencial del Item (DIF, por sus siglas en inglés) y equiparación de puntajes (equating) no solo revisten gran importancia conceptual desde el enfoque de validez, sino también desde las dimensiones de equidad y la justicia para las decisiones que se basan en los resultados de las pruebas. En el caso del Funcionamiento Diferencial del Item se busca identificar reactivos o preguntas que, en vez de reflejar diferencias reales en el constructo que se desea medir, reflejen, erróneamente, otros factores no asociados al constructo y que interactúan con características sociodemográficas del examinado, como sexo, lugar de procedencia y nivel socioeconómico (Penfield & Camilli, 2007). Por su parte, la equiparación de puntajes, se refiere a métodos para hacer comparables en dificultad resultados de formas diferentes de una misma prueba. En este caso las formas diferentes de una misma prueba están compuestas por exámenes con ítems diferentes pero que intentan medir el mismo constructo. (Dorans & Holland, 2000).

Características de la prueba PISA que la hacen una evaluación única en su género

Las principales características de PISA que la hacen una evaluación única en su género se detallan a continuación (OCDE-Santillana, 2011, p. 18-19).

- Se concentra en la capacidad del estudiante de utilizar sus habilidades y conocimientos para enfrentarse a los retos de la vida real. Es decir, se preocupa por tratar de determinar lo que el estudiante sabe y puede hacer con lo que aprendió en el sistema escolar y no si domina y reproduce contenidos curriculares específicos. Se trata de una prueba acurricular.
- Maneja un concepto innovador de “competencia”, entendida como la capacidad del estudiante de aplicar sus conocimientos y habilidades a determinadas áreas

disciplinarias y de analizar, razonar y comunicarse con eficacia cuando plantean, interpretan y resuelven problemas en diversas situaciones.

- Se orienta hacia políticas educativas, lo que conecta los datos sobre los resultados educativos de los estudiantes con los datos sobre sus características y sobre factores claves que inciden en su aprendizaje dentro y fuera de centro escolar. Esto le permite centrarse en las diferencias en los modelos de rendimiento e identificar las características de los estudiantes, los centros y los sistemas educativos con niveles de rendimiento altos.
- Da relevancia al aprendizaje a lo largo de la vida, pues no se limita a evaluar las competencias estudiantiles en determinados campos disciplinarios, sino que también busca conocer sus motivaciones para aprender, sus opiniones sobre sí mismos y sus estrategias de aprendizaje.
- Se aplica con regularidad, lo que permite a los países hacer un seguimiento de su progreso en el logro de los objetivos claves de su sistema educativo.
- Tiene amplia cobertura geográfica y metodología colaboradora de elaboración y aplicación, ya que los expertos de los países y economías participantes interactúan y cooperan con los expertos internacionales de PISA y entre sí.
- Aplica mecanismos de control de calidad muy estrictos en la concepción, traducción y recolección de muestras de textos, de datos y en el diseño de la prueba, lo cual hace que los resultados de PISA tengan validez y confiabilidad.
- Se han demostrado la relevancia y el valor predictivo de las mediciones de conocimientos y habilidades efectuadas por PISA. Estudios longitudinales llevados a cabo en Australia, Canadá y Suiza muestran que hay una estrecha relación entre el rendimiento en lectura evaluado por PISA a los 15 años de edad y posteriores resultados educativos y éxitos en el mundo laboral. Esto ocurre porque los niveles de competencia lectora son predictores más fiables del bienestar económico y social que la cantidad de educación medida en años de escolarización. Lo que importa es la calidad de los resultados del aprendizaje y no la duración de la escolaridad. (OCDE-Santillana, 2011; OCDE, 2010f; OCDE, 2010g)
- A las características anteriores, citadas por el mismo PISA, debemos agregar de nuestra cuenta que en un contexto como el de nuestro país, nuestra participación en este programa no solo proporciona una medida de desempeño de alta confiabilidad y validez, que permite conocer el perfil de nuestros estudiantes de 15 años en el sistema de educación formal en cuanto a conocimientos y habilidades necesarios para enfrentar el mundo actual, sino que también nos permite utilizar el rico marco de referencia conceptual y experiencia acumulada de este programa para realizar estudios de factores asociados al rendimiento que puedan fundamentar acciones de política educativa.

- De nuevo reiteramos que PISA, como prueba estandarizada de alta calidad técnica, utiliza modelos de medición y análisis “de punta” para generar evidencias sólidas del grado de validez y confiabilidad de sus resultados, propiedades que posibilitan las interpretaciones sustantivas para los significados los puntajes numéricos obtenidos a partir de la administración de las pruebas. (OECD, 2012). Específicamente, se puede mencionar el uso de los modelos de Rasch para lograr inferencias criterio válidas acerca de los niveles de desempeño de los estudiantes, es decir, para establecer qué puede y qué no puede lograr un examinado particular, de acuerdo con los puntajes obtenidos. También merece mencionarse el uso de métodos para detectar funcionamiento diferencial de los ítems (DIF) en el contexto específico de un país particular y la eliminación de estos ítems de los cuadernillos de examen que se administran en ese contexto. Igualmente destaca el empleo de complejos modelos de “equating” (equiparación de puntajes) para garantizar que las diferencias de dificultad entre los diferentes cuadernillos o fórmulas de examen y el diferente muestreo de contenidos no afecten las estimaciones del nivel del constructo o calificación de los examinados.
- En este sentido hay que tener presente también que las calificaciones otorgadas por los profesores poseen un valor relativo como medida de rendimiento, ya que no existe un criterio estandarizado para todos los centros educativos, todos los cursos y todo el cuerpo docente. Igualmente, las pruebas estandarizadas de contenidos curriculares se enfocan más en la medición de los conocimientos que se supone deben ser el producto de un programa formal de estudio, sin atender de manera directa cómo se usan estos conocimientos en el contexto de la vida cotidiana y cómo ayudan a enfrentar apropiadamente los retos del mundo moderno.

Los informes de resultados generales de PISA 2009

A los fines de dar a conocer los resultados de sus pruebas, PISA produce una serie de informes en varios volúmenes. Es importante aclarar que nos referimos aquí a los resultados generales de las pruebas y los cuestionarios aplicados en cada edición de PISA. Adicionalmente, PISA produce y edita a través de la OCDE otra gran cantidad y variedad de análisis e investigaciones asociadas a las pruebas, sea sobre países o temas específicos.

Los resultados de PISA 2009 se presentaron en seis volúmenes, lo cual muestra un incremento considerable en comparación con la edición anterior de la prueba (2006), que consistieron solamente en dos volúmenes. Además se cuenta con un volumen adicional que presenta los resultados de PISA 2009+ (Walker, 2011).

Seguidamente se resumen los contenidos de los seis volúmenes de resultados de PISA 2009. El primero, disponible en español –*Lo que los estudiantes saben y pueden hacer* (OCDE-Santillana, 2011)–, se concentra en un análisis general de los resultados obtenidos por los estudiantes en las tres áreas evaluadas: competencia lectora, matemática y científica. Como se dijo, el ámbito de la lectura recibe especial atención por ser el énfasis de la evaluación del 2009. El segundo volumen, disponible a la fecha solo

en inglés –*Overcoming Social Background: Equity in Learning Opportunities and Outcome* (OCDE, 2010a)— analiza la relación entre la situación socioeconómica del alumnado y su rendimiento en competencia lectora. Este es un indicador de la equidad de los sistemas educativos y su eficacia para reducir desigualdades de origen.

El tercer volumen –*Learning to Learn: Student Engagement, Strategies and Practices* (OCDE, 2010b)— ofrece información que hasta ahora ha sido la menos explorada de los resultados de PISA: las actitudes y hábitos de los estudiantes hacia la lectura y las estrategias que utilizan. El cuarto volumen –*What Makes a School Successful? Resources, Policies and Practices* (OCDE, 2010c) – se centra en el estudio de los centros escolares y su relación con el desempeño en lectura de sus estudiantes, partiendo del supuesto que lo que pasa en el centro incide de manera directa en el aprendizaje. Se añaden nuevos aspectos que no se habían analizado en ediciones anteriores de PISA, como las expectativas de los alumnos hacia el centro, el clima de aula, las relaciones entre alumnos y docentes o la disciplina escolar. Otros factores estudiados en relación a los resultados en lectura son la selección y la agrupación de los estudiantes, la autonomía de los centros, o los recursos financieros que se destinan a ellos.

El quinto volumen –: *Learning Trends. Changes in Student Performance Since 2000* (OCDE, 2010d)— analiza los progresos que han hecho los países participantes desde el inicio de PISA en cuanto a elevar el desempeño estudiantil y mejorar la equidad en la distribución de las oportunidades de aprendizaje. Los cambios en los logros a lo largo del tiempo demuestran que el rendimiento de un país no está predeterminado; los resultados de la educación son mejorables en todos los sistemas educativos sin se dan las condiciones necesarias para ello. (Castejón, 2011) El sexto volumen –*Students On Line. Digital Technologies and Performance* (OCDE, 2011b)— examina el uso que hacen los estudiantes de las tecnologías de la información para aprender.

Algunas limitaciones de PISA

- Existe una crítica al contenido sustantivo de los ítems de las pruebas PISA en el sentido de que incluyen contextos muy centrados en la cultura occidental y de un país desarrollado. Quizá esta crítica se fundamenta especialmente en el hecho de que muchos de los ejemplos de reactivos que se liberan para conocimiento del público se refieren a este tipo de contenidos. Sin embargo, PISA al incluir como participantes a países tan diversos establece protocolos para evitar que ítems con evidente sesgo cultural o socioeconómico se incluyan en los cuadernillos de examen de la aplicación particular de cada país. Se intenta así minimizar la posibilidad de sesgos importantes en las respuestas debido a efectos de contexto, cultura o vocabulario. Por esta razón todos los reactivos de los cuadernillos que se aplicarán en un país deben ser juzgados previamente por un equipo del país en cuanto a su validez para el contexto educativo local. En Costa Rica esta revisión estuvo a cargo de personal técnico del Departamento de Evaluación Académica y Certificación de la Dirección de Gestión y Evaluación de la Calidad del Ministerio de Educación Pública, y, de acuerdo con su valoración, se realizaron los cambios o modificaciones

necesarias en aquellos reactivos donde se sospechó alguna posibilidad de sesgo por cultura, contexto o vocabulario. (P. Mena, comunicación personal, junio 2012).

- Otro cuestionamiento relacionado con el anterior se refiere a que las pruebas PISA tienen más precisión y están mejor adaptadas a los niveles de competencia de los estudiantes en los países con mayor desarrollo. Con respecto a este punto es necesario indicar que en años recientes este programa ha realizado un esfuerzo por superar esta crítica, y a partir de la aplicación del año 2009 se ensamblan cuadernillos o formas de examen de menor dificultad para ser aplicados en aquellos países participantes que obtienen puntajes promedio relativamente más bajos. Esto se realiza con el propósito de representar con mayor precisión los niveles de menor puntaje en la escala de desempeños. (OECD, 2012). Los procesos de equiparación de puntajes (equating) permiten ajustar o controlar el efecto de estas diferencias en dificultad, de manera que los puntajes sigan siendo comparables.
- En sus cuestionarios de contexto dirigidos a estudiantes y directores, PISA incluye variables que, de acuerdo con estudios internacionales y sus propias investigaciones, tienen relevancia general para la predicción del rendimiento en las pruebas, pero precisamente por este carácter, estos instrumentos no permiten medir factores idiosincráticos del contexto del país particular, que podrían estar asociados al rendimiento de sus estudiantes.
- PISA, por su naturaleza acurricular, solo se enfoca en tres áreas, Ciencias, Matemáticas y Comprensión Lectora, no pudiéndose generar ninguna evidencia de desempeño para el resto de las materias de la enseñanza formal.
- El diseño muestral utilizado por PISA no permite medir y analizar variables y dimensiones del docente y de la dinámica del aula que pueden tener relevancia en la predicción del rendimiento. Esto sucede porque en cada centro educativo seleccionado se toma una muestra al azar de 30 estudiantes de 15 años, que pueden pertenecer a diferentes secciones, grupos o niveles. En otros programas internacionales como TIMSS y TERCE, así como en los mismos estudios de factores asociados a los resultados de las pruebas diagnósticas que se realizan en el Ministerio de Educación Pública de nuestro país, sí se utiliza un muestreo de secciones o grupos de estudiantes con sus correspondientes profesores y así se pueden medir e incorporar en los modelos multinivel dimensiones pedagógicas y características del docente y de la clase para examinar su poder predictivo para las notas en las pruebas. Si bien es necesario puntualizar que desde la perspectiva de PISA esta no es una limitación importante, principalmente porque los resultados de los estudiantes, particularmente en el nivel secundario, no son el producto de un solo docente. El estudiante ha tenido varios profesores de la materia antes del actual. ¿En qué grado uno puede de manera legítima asociar al profesor actual con los resultados de los estudiantes que son el producto de muchos años de aprendizaje? La pregunta no tiene una respuesta sencilla. (M. Walker, comunicación personal, 31 de enero 2013).

- Quizá el mayor peligro de una interpretación no precisa de los resultados de PISA en el caso costarricense tiene que ver con el hecho de que, como se mencionó al principio y como se expone de manera más detallada en la sección “Representatividad de la muestra de PISA en Costa Rica”. En esta aplicación del año 2009 la muestra de PISA es representativa solamente para un 53% de los jóvenes de 15 años, dado que los muchachos que no están en el sistema educativo no se incluyen en el marco muestral, ni tampoco los que están en el sistema denominado “de Educación Abierta”, ni los que, estando en el sistema formal de educación, requieren de algún tipo de adecuación en la aplicación de los exámenes. Por tanto, la opinión pública y los responsables de las políticas educativas deben tener muy claro que si se incluyeran estas poblaciones en PISA, las cuales conjuntamente representan casi la mitad de los jóvenes de 15 años, sería muy probable que los desempeños en estas pruebas disminuirían considerablemente, tal como se ilustra en la mencionada sección, utilizando un modelo de simulación.

Marco general de referencia: algunos determinantes del rendimiento académico

Cada año se publican, a nivel internacional, gran cantidad de artículos y reportes en relación con el estudio del rendimiento académico y el abandono escolar, ámbitos que han sido una preocupación constante en el campo de la investigación educativa. El énfasis se concentra actualmente en encontrar relaciones causales entre el desempeño académico, medido por medio de indicadores homogéneos, como pruebas estandarizadas, y variables que puedan ser objeto de intervenciones programáticas (por ejemplo, uso de materiales didácticos o estrategias pedagógicas).

Este tipo de investigación es clave, si se pretende brindar a los tomadores de decisiones y a la opinión pública los instrumentos necesarios para implementar cambios a nivel institucional o programático, cambios que, de acuerdo con los hallazgos investigativos, produzcan efectivamente, impactos positivos en las variables educativas.

Por ejemplo, un resultado de investigación internacionalmente aceptado es que existe una asociación significativa entre el nivel socioeconómico del estudiantado y su desempeño académico. Sin embargo, a nivel de una intervención, es muy difícil concebir que una institución educativa pueda hacer algo para cambiar la primera condición. Siempre es importante la medición de este tipo de relaciones, con el propósito de conocer científicamente el fenómeno, pero es poco lo que se puede recomendar a partir de tales resultados de investigación. Es más, desde una perspectiva causal el nivel socioeconómico en sí mismo no brinda una explicación exhaustiva en torno al por qué se da esta relación observada, sino que esta variable es probablemente más bien un proxy de un conjunto mayor de dimensiones relacionadas con oportunidades educativas. Por el contrario, si se encuentra una relación significativa entre ciertas características del centro educativo o la metodología empleada por el docente y el rendimiento, es relativamente más sencillo sugerir entonces acciones de política educativa que se refieran a esas dimensiones institucionales o pedagógicas que pueden ser objeto directo de intervención. De igual forma, a nivel teórico tiene mayor

sentido plantear los procesos causales que relacionan estas variables con el rendimiento académico.

Así, el rendimiento estudiantil es un fenómeno multicausal, provocado por conjuntos de variables en diversas dimensiones que usualmente se correlacionan entre sí, y entre todas remiten a marcos de referencia teóricos que deben utilizarse para comprender cómo sucede el fenómeno y dónde intervenir, si se desean generar cambios para lograr mejoras en el rendimiento. Todos estas variables representan elementos que considerados independientemente generan efectos relativamente menores, pero que considerados conjuntamente generan las sinergias necesarias para provocar logros sustanciales en metas educativas (Zamora, 2012).

A continuación se presentan una serie de argumentos basados en una revisión de literatura que describen algunos de esos factores que intervienen en el desempeño escolar y que consecuentemente también podrían explicar o predecir el rendimiento de nuestros estudiantes en pruebas educativas como PISA, como las pruebas nacionales de Bachillerato o como las pruebas nacionales de diagnóstico de sexto y noveno año.

Factores de contexto del país

Los **factores culturales, sociales, políticos y económicos** del contexto afectan al sistema educativo de un país. Las decisiones sobre la estructura de la organización educativa, recursos, instalaciones, capacitación docente y el currículo dependen a menudo de estos factores. Según el marco conceptual en el que se basan las pruebas TIMSS, el éxito que tiene un país en la enseñanza de las matemáticas y las ciencias depende del valor que le dé la sociedad a estas materias, los recursos disponibles y los mecanismos que se ofrecen para proporcionar contextos de aprendizaje efectivos. (Mullis et al, 2011).

Las características de la población de un país y la economía nacional pueden tener un tremendo impacto en la facilidad o dificultad para proporcionar contextos eficaces para el aprendizaje. El tamaño geográfico de un país puede crear dificultades en la entrega uniforme de un plan de estudios. Los países con una población grande y diversa y con pocos recursos humanos, se enfrentan, en general, a mayores desafíos que los que tienen circunstancias más favorables (en Bos, Schwippert, & Stubbe, 2007; Gradstein & Schiff, 2004; Kirsch, Braun, Yamamoto, & Sum, 2007; Taylor & Vinjevd, 2000; Trong, 2009, citado por Mullis et al, 2011).

Características del sistema educativo: Otro aspecto a considerar se refiere al sistema educativo, al respecto autores como He-chuan Sun (citado por Blanco et al, 2008), concluyó, a partir de un estudio realizado en 8 países, que todos ellos tenían objetivos y metas de aprendizaje descritos en el currículo nacional o en los lineamientos curriculares, pero que la diferencia radicaba en que algunos de ellos tenían “prolongaciones” hacia la práctica y otros no. Según la evidencia, la parte más activa de estos objetivos es la que está en los libros de texto. El estudio revela que los objetivos

nacionales no influyen en el mejoramiento escolar, sino más bien las metas académicas más específicas para cada materia, establecidas a nivel nacional. Para que se produzca un efecto positivo las acciones curriculares deben estar respaldadas por un sistema nacional de exámenes, la evaluación externa (sistema de inspección), sistemas de feedback y refuerzos (Blanco et al, 2008).

Expectativas altas para los alumnos: De acuerdo con el informe de McKinsey (Barber & Mourshed, 2008) todos los sistemas con alto desempeño y rápida mejora cuentan con planes de estudios que fijan expectativas claras y altas con relación a lo que los alumnos deben lograr. Por ejemplo, la provincia de Alberta en Canadá fija altas expectativas académicas y luego participa en exámenes internacionales como PISA y TIMSS para tener una referencia de sus estándares. Finlandia reformó sus planes de estudio en 1992, sustituyendo el rígido plan de estudios nacional por uno basado en objetivos para todos los estudiantes (Barber & Mourshed, 2008).

Características del estudiante

Nivel socioeconómico: Según las evaluaciones de PISA en los países donde existe mayor nivel de repetición, el nivel socioeconómico suele ser menor, lo que sugiere que los grupos de menor nivel socioeconómico son más propensos a ser afectados negativamente por el grado de repetición. En muchos países, los centros educativos que compiten por los estudiantes tienden a tener un mayor rendimiento, pero esto es menudo explicado por el mayor nivel socioeconómico de los estudiantes que asisten a estos centros. Esto sucede porque los padres con un nivel socio económico alto, al momento de elegir a cuál centro enviar a sus hijos, tienden a considerar más el rendimiento académico de los centros educativos que los padres de un nivel socioeconómico bajo. En general, los padres citan los logros académicos como una de los factores más importantes a la hora de elegir un centro educativo para sus hijos (OECD, 2010). Aquellos padres de familia que cuentan con mayor nivel socioeconómico tienen mayor probabilidad de considerar esta variable como muy importante, comparados con los padres de menor nivel socioeconómico. Posiblemente la diferencia tenga su origen en que los padres con situación socioeconómica desfavorable tengan menos opciones para elegir los centros educativos, debido a sus limitaciones financieras.

Contexto familiar: Los resultados de TIMSS sirven para ilustrar cómo el contexto familiar (medido por cantidad de libros en el hogar, disponibilidad de una mesa de estudio, presencia de una computadora, educación de cada uno de los padres, número de padres en el hogar, número de familiares que viven en el hogar y las posesiones en el hogar) y el rendimiento de los estudiantes están estrechamente entrelazados. El contexto familiar de los estudiantes y la influencia de las comunidades son predictores potentes del logro en ciencias y matemáticas. Esta es una relación que se manifiesta y persiste a través de contextos internacionales. (Martin et al, 2000).

El contexto familiar incluye tanto el nivel socioeconómico de los padres como el educativo, algunas investigaciones demuestran una relación positiva entre los logros estudiantiles y los indicadores de estatus socioeconómico, como el nivel educativo de los padres y su ocupación (en Bradley y Corwyn, 2002; Haveman y Wolfe, 2008; Willms, 2006, citado por Mullis et al, 2011).

El nivel educativo de la madre, por ejemplo, ha llamado la atención en el estudio de Mella y Ortiz (1999) al determinar que la madre es el agente socializador fundamental, pues desde el comienzo de la vida se comunica con el niño transmitiéndole su nivel cultural por medio del lenguaje y la relación afectiva. Además, en forma sistemática, entrega las normas y valores que serán los referentes de la conducta del niño. Menos años de estudio de la madre se asociarían con menor logro escolar por parte de los niños. En el análisis realizado por estos autores, el conjunto de las dos variables asociadas a la madre, su nivel de escolaridad y sus expectativas, logra un poderoso nivel explicativo (0.37) reflejado en el coeficiente estandarizado de regresión, llegando a ser mucho más importante que el nivel de ingresos familiares y que la educación del jefe de hogar (Mella y Ortiz, 1999).

Autoconfianza y motivación: Las características individuales del estudiante son otra causa de diferenciación en su desempeño. Factores aparentemente “intangibles” pueden afectar el aprendizaje de los alumnos. El interés, el entretenimiento durante la realización de la tarea, así como la percepción que tienen los alumnos de sus propias posibilidades para enfrentarse a determinados problemas, pueden llegar a ser más relevantes, incluso, que el disponer de ciertas condiciones materiales. PISA 2006 recoge, entre otros, datos sobre la actitud y el involucramiento de los alumnos respecto a la ciencia, los cuales, analizados con modelos de Ecuaciones Estructurales y con Dual Scaling permiten valorar el peso de los distintos “intangibles” en el rendimiento académico. Se comprueba que la confianza, motivación y el interés contribuyen al rendimiento de un estudiante. La confianza en las propias destrezas científicas (self-efficacy) y el disfrutar realizando tareas relacionadas con la ciencia, parecen los mejores indicadores de un buen rendimiento en las competencias científicas (Clavel & Balibrea, 2010).

Entre los países de la OCDE, el 31% de los estudiantes que provienen de entornos poco favorables son “resilientes”, lo cual significa que están entre los que, internacionalmente, tienen mejor desempeño de todos los estudiantes de entorno sociocultural similar. Una diferencia clave entre los alumnos de entornos poco favorables que son resilientes y los que no lo son es que los primeros asisten más y regularmente a clases en el centro de enseñanza. Resultados de PISA muestran que, cuanta más auto confianza y motivación tengan los alumnos, mayores serán sus probabilidades de ser resilientes. (OECD, 2011). Es decir, los alumnos que se encuentran en situaciones socioeconómicas desfavorables no están condenados a perpetuar un ciclo intergeneracional de malos resultados académicos y con ello malas perspectivas de trabajo y pobreza, si asisten a centros educativos que imparten más clases en forma regular. Entre los alumnos de entornos poco favorables, el tiempo dedicado al aprendizaje en el centro educativo es uno de los predictores más poderosos para identificar a los estudiantes van a desempeñarse mejor que sus pares.

El alumno resiliente típico dedica más tiempo a estudiar Ciencias en el centro educativo (en promedio entre una y dos horas más por semana) que el alumno típico de entornos poco favorables y que logra bajos resultados, esto sucede tanto en los países de la OCDE como en los países y economías asociados. Por ejemplo, en Francia, Alemania y Holanda los alumnos resilientes dedican por lo menos una hora y 45 minutos más por semana a la clases de Ciencias que los alumnos de entornos desfavorables que logran bajos resultados (OECD, 2011).

Otro factor asociado a la resiliencia es la **confianza** que tienen los alumnos en sus propias habilidades académicas. Resultados de PISA muestran que, cuanto más confianza en sí mismos tengan los alumnos, mayor será su posibilidad de ser resilientes. La aplicación PISA 2006 reveló que más del 50% de los alumnos resilientes en los países de la OCDE opinaron que el aprendizaje de temas avanzados de ciencias sería fácil para ellos, mientras que aproximadamente un 40% de los alumnos de entornos poco favorables y con bajo desempeño pensaba de igual manera. Aproximadamente el 75% de los estudiantes resilientes creía que puede contestar correctamente a preguntas de examen sobre temas de Ciencias, mientras que aproximadamente solo un 50% de los alumnos de entornos poco favorables con mal rendimiento compartía esta creencia. La **motivación**, en particular la motivación que surge de un impulso personal e interno, en lugar de la motivación que surge de un estímulo externo, está también asociada con la característica de resiliencia para los alumnos en muchos países, aunque esta relación es más débil. (OECD, 2011).

Habilidades cognitivas y no cognitivas al nacer: De acuerdo al estudio realizado en Alemania por el Centro de Investigación Económica y el Instituto Central de Salud Mental, se apoya la evidencia de que tanto las habilidades cognitivas y no cognitivas parecen contribuir al éxito económico futuro de los estudiantes (en Borghans et al. 2008, Heckman 2008, citado por Coneus, Laucht & Reub (2010). Los resultados indican que las habilidades cognitivas y mentales son importantes predictores del éxito en la escuela, y la predicción aumenta con la edad. Desde los tres meses de edad, las habilidades cognitivas predicen de manera significativa el alcance de los grados escolares. A partir de los dos años en adelante, los niños con altas capacidades mentales tienden a desempeñarse mucho mejor en la escuela. Si se agrega al coeficiente las habilidades emocionales logradas a la edad de 11 años, se mejoran las calificaciones en la escuela a casi la mitad de un grado. Por lo tanto, se concluye que las habilidades cognitivas al nacer son un indicador importante para el éxito escolar y la relación aumenta con la edad (Coneus, Laucht & Reub, 2010).

Educación preescolar: Los países que tienen mayor cobertura de su sistema preescolar obtienen mejores resultados pues sus estudiantes tienden a desempeñarse mejor que aquellos que provienen de países con menor cobertura en preescolar. Los sistemas educativos con una mayor proporción de estudiantes con educación preescolar tienden a obtener mejores resultados académicos (OECD, 2010). En más de la mitad de todos los países de la OCDE, más del 94% de los estudiantes con 15 años de edad, informaron que habían asistido a preescolar por lo menos durante algún tiempo. Los mejores resultados se observan en aquellos sistemas escolares en los que

la educación preescolar dura más tiempo, donde hay grupos más pequeños de alumnos por docente y donde hay un mayor gasto público por alumno en ese nivel de educación.

Los sistemas educativos de alto desempeño ponen un fuerte énfasis en aritmética y lengua durante los primeros años de escolaridad, basados, en parte, en investigaciones que demuestran que la habilidad temprana en capacidades fundamentales está íntimamente relacionada con un amplio espectro de resultados futuros: por ejemplo, un importante estudio longitudinal en el Reino Unido determinó que los resultados de las pruebas de lengua y aritmética a los siete años resultaban determinantes para los logros obtenidos a los 37, aun después de ajustar estadísticamente por nivel socioeconómico (Barber & Mourshed, 2008).

Asistencia: Investigaciones demuestran que la asistencia consistente de estudiantes y profesores a clases se relaciona con un mayor rendimiento. Si los estudiantes no asisten a la escuela regularmente se reduce radicalmente su oportunidad de aprender. La literatura ha demostrado que los estudiantes de las escuelas donde los directores reportan problemas de asistencia obtienen un menor rendimiento en las pruebas TIMSS (Mullis et al, 2011).

Género: Las investigaciones también han apuntado a que la diferencia de rendimiento entre los niños y las niñas en matemáticas y ciencias suele ser menor que las diferencias asociadas a factores de origen, es decir, mientras que durante décadas ha habido una preocupación acerca del rezago de las niñas en matemáticas y ciencias, en la actualidad la mayoría de investigaciones no lo consideran un factor determinante (en Coley, 2001; McGraw, Lubienski, y Strutchens, 2006, citado por Mullis et al, 2011). TIMSS ha demostrado que no hay gran diferencia en el promedio global de matemáticas y ciencias entre niños y niñas de los países participantes, aunque la situación varía de país a país. Sin embargo, con datos de la OCDE del Programa para la Evaluación Internacional de Estudiantes (PISA), los investigadores han encontrado diferencias en la media y la varianza de las puntuaciones de las pruebas de matemáticas y lectura, indicando que las niñas generalmente obtienen más puntos que los varones en lectura, mientras que los muchachos logran mejores resultados que las niñas en matemática. (Machin and Pekkarinen 2009)

Efecto inmigrante y pertenecer a una minoría: Desde la aparición del conocido informe Coleman (1966), se han elaborado para EE.UU. diversos trabajos que muestran diferencias de rendimiento entre alumnos blancos nativos y alumnos de otras minorías étnicas, como latinos, negros o asiáticos. Estas diferencias afectan tanto a los inmigrantes como a aquellos que aunque no tiene condición de inmigrantes pertenecen a alguna minoría étnica. Diversos estudios han intentado explicar estas diferencias a través de variables relativas a las características individuales de estos alumnos, ya sean de carácter socioeconómico o sociológicas (en Portes y Rumbaut, 1990; Rong y Grant, 1992; Kao y Tienda, 1995; Bankston y Zhou, 2002; Chiswick y Debburman, 2004, citado Zancajo & Franquesa, 2010).

En un estudio realizado por la Universidad Autónoma de Barcelona sobre la descomposición del efecto inmigrante en el rendimiento académico se detalla que la

comparación internacional de PISA - 2006 muestra como los alumnos de origen inmigrante tienen un diferencial de rendimiento respecto a los alumnos nativos. A través de la comparación internacional que permiten estos datos de PISA se observa como el diferencial de rendimiento, entre alumnos nativos e inmigrantes, varía según el país analizado (OCDE, 2006). En las pruebas PISA 2006, para el área de ciencias, la diferencia de rendimiento entre nativos e inmigrantes de primera generación oscilaba entre -108 puntos en Bélgica y -12 puntos en Australia. Aunque estas diferencias se reducen si se examinan las puntuaciones de los alumnos inmigrantes de segunda generación, la brecha se mantiene para la mayoría de países analizados. Además, alumnos inmigrantes que hablan en el hogar el mismo idioma que se utiliza en la escuela tienen mejor rendimiento que aquellos que utilizan lenguas diferentes, consecuentemente las políticas de integración tienen un papel determinante (en Entorf y Minoiu, 2005, citado por Zancajo & Franquesa, 2010).

Otros estudios revelan que si bien las condiciones culturales y sociales son importantes, las características socioeconómicas de los diferentes grupos de inmigrantes en cada país anteceden a las primeras para explicar su efecto sobre el rendimiento de los alumnos inmigrantes, aunque también se consideran influyentes otras variables como el idioma materno o los años de residencia en el país (en Schnepf, 2004; Marks, 2005, citado Zancajo & Franquesa, 2010)

Salud: Las investigaciones realizadas muestran que los niños sanos y bien nutridos aprenden más, pues los menores que comen mejor a lo largo del día, rinden más en la escuela que quienes tienen dietas menos saludables. En un estudio publicado por la Escuela de Salud Americana los estudiantes con dietas de baja calidad tuvieron significativamente más probabilidad de obtener un mal desempeño en sus calificaciones. Las niñas obtuvieron mejores resultados que los varones al igual que los hijos de familias socioeconómicamente favorecidas. Los niños que asisten a mejores escuelas y que viven en los barrios ricos también obtuvieron mejores resultados. El estudio contempló a casi 5 mil niños de los cuales los que presentaban mayor calidad de la dieta fueron significativamente menos propensos a fallar en las pruebas de literatura, sus probabilidades de fracasar eran entre un 26% y 41% menores. (Florence 2008). Estos resultados, explican los expertos, se deben a que una alimentación adecuada y variada durante el día provee glucosa, es decir, la energía necesaria para que exista una buena actividad cerebral. También los nutrientes como ciertas vitaminas y minerales obtenidos a partir del consumo de frutas y verduras mejoran la disponibilidad de la acetilcolina, un neurotransmisor que se requiere para los procesos de memorización. (Asús, 2013).

Características del centro educativo

Eficacia escolar: Un estudio pionero que ha mostrado el peso relativamente significativo del contexto social ha sido el así llamado Informe Coleman (1966), realizado en Estados Unidos a mediados de la década de los sesenta. El tema central de este estudio se refirió a si los insumos escolares pueden, por sí mismos, tener un efecto específico sobre el rendimiento o el éxito escolar de los alumnos, más allá de los factores socioeconómicos. El Informe Coleman concluyó que los factores o insumos escolares tenían, al parecer, un efecto poco significativo sobre las diferencias en el desempeño escolar. Estos resultados implicaron un cuestionamiento a la escuela como agente democratizador (Mella y Ortiz, 1999).

Sin embargo, es difícil aceptar que la intervención del sistema escolar no deja huellas en los estudiantes. En un meta-análisis publicado en la Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, se detalla que 12 de cientos de estudios realizados sobre los efectos de las escuelas obtienen que, en promedio un 18.49% de la varianza en el desempeño escolar se explica por el efecto de la escuela (Carballo, 2005).

Observaciones cualitativas y de expertos señalan que el impacto del sistema escolar no deja de ser significativo. En Latinoamérica se han realizado varios estudios sobre eficacia escolar: Lockheed y Bruns en Brasil trabajaron con un modelo de niveles múltiples que permitió explorar el porcentaje de varianza explicada en los resultados de la escuela secundaria brasileña, el cual se atribuyó a: i. la modalidad del centro educativo; ii. las diferencias en los resultados en matemática y portugués que se encontraron entre centros educativos, y, iii. las diferencias en la relación entre rendimiento y situación socioeconómica de los estudiantes. De estas tres variables, se encontró que una parte significativa de la varianza de los resultados se debe a la modalidad del centro educativo, ya fuera este público o privado. (Mella y Ortiz, 1999).

Algunos estudios reportan entre un 8% y un 15% de varianza explicada para los resultados escolares a partir de factores del sistema educativo. La mayor variación entre las escuelas se debe a diferencias en la dinámica del salón de clases (Scheerens et al., 1989). Además, cabe señalar que los efectos de una escuela pueden variar muy rápidamente, incluso entre dos y tres años (en Nuttall et al., 1989, citado por Mella y Ortiz, 1999).

Javier Murillo (Murillo, 2004) investigador de la Universidad Autónoma de Madrid, desarrolló un marco para comprender la mejora de la eficacia escolar a partir de la revisión de teorías educativas y de experiencias de mejora de eficacia escolar llevadas a cabo en 8 países de Europa, obteniendo el marco descriptivo que se presenta a continuación.

Cuadro 2
Factores de mejora de la eficiencia escolar procedentes del análisis de casos

Factores de mejora de la eficacia escolar procedentes del análisis de casos

	Países							
	PB	Fin	Bel	RU	Esp	Por	It	Gre
Contexto								
Agentes externos implicados en programas de mejora	+		+	+	x	+	+	+
Presión externa para comenzar la mejora	+		+	+	+	+	+	
Evaluación externa de las escuelas	+		+	+			+	+
Competencia entre centros								
Descentralización de las decisiones (contenido, práctica de enseñanza)			-	-				+
Escuela								
Actitud positiva en relación con el cambio	+	+			+		+	
Cultura escolar, valores compartidos, visión, misión	+	+			+			
Organización escolar que facilita la mejora (tiempo, etc.)					+	+	+	+
Liderazgo del director (u otros miembros de la plantilla)	+	+		+	+			
Inestabilidad de la plantilla	-				-	-		-
Evaluación interna (valoración de estudiantes y profesores)	+	+						+
Establecimiento de objetivos (resultados y/o obj. intermedios)	+	+		+				
Implicación de los padres/comunidad en los programas de mejora					x	+	+	
Planificación adecuada del proceso de mejora		+			+			
Mejora incluida en el desarrollo total de la escuela	+			+				
Complejidad/comprensividad del programa de mejora	-				+			
Ciclo de mejora auto-regulado					+			
Preparación para el cambio/abordando primero las cuestiones visibles				+				
Participación del alumnado en los esfuerzos de mejora							+	
Aula/profesor								
Motivación e implicación/participación del profesorado en procesos y decisiones		+			+		+	+
Colaboración del profesorado (en la escuela, a través de escuelas)		+				+	+	
Feedback del comportamiento del profesorado	+	+						
Formación de profesorado					+			
Operación de elementos esenciales de currículos/innovaciones						+		
+: influencia positiva	PB	Países Bajos			Esp	España		
-: influencia negativa	Fin	Finlandia			Por	Portugal		
x: sin influencia	Bel	Bélgica			It	Italia		
	RU	Reino Unido			Gre	Grecia		

Fuente: Murillo (2004)

Gestión del tiempo: De acuerdo con Javier Murillo, el grado de aprendizaje del alumno está directamente relacionado con la cantidad de tiempo que está implicado en actividades de aprendizaje. Así, un aula eficaz será aquella que realice una buena gestión del tiempo, de tal forma que se maximice el tiempo de aprendizaje de los alumnos. En un aula eficaz, el docente optimiza el tiempo de las clases para que esté

lleno de oportunidades de aprendizaje para los alumnos. Ello implica disminuir el tiempo dedicado a las rutinas, a la organización de la clase o a poner orden. Cuanto menos frecuente y más breves sean las interrupciones, más oportunidades el alumno tendrá para aprender (Blanco et al, 2008).

Diferenciación entre estudiantes: De acuerdo con la OECD (2010c) lo que sucede en el centro educativo tiene un impacto directo en el aprendizaje de los estudiantes y el desempeño de los centros educativos está influido por los recursos, las políticas y las prácticas que definen, a nivel superior, los sistemas de educación en un país. Los sistemas escolares exitosos proporcionan a todos los estudiantes, independientemente de sus antecedentes socioeconómicos, las mismas oportunidades para aprender. Los sistemas escolares que diferencian a los estudiantes de acuerdo a la forma en que los asignan en las escuelas, clases y grados muestran resultados menos equitativos.

Educación privada o pública: Las centros educativos con mejores recursos tienden a lograr mejores resultados académicos en la medida en que también tienden a tener estudiantes con más altos niveles socio-económicos. Las escuelas deben contar con los recursos necesarios mínimos para ofrecer una enseñanza efectiva, aunque recursos adicionales pueden producir solo una pequeña diferencia en los resultados. Luego de controlar los perfiles socio-económicos y demográficos de los estudiantes y de las escuelas, los estudiantes de países de la OCDE que asisten a instituciones privadas muestran un rendimiento similar al de los estudiantes matriculados en las escuelas públicas. (OECD, 2011b)

Los estudiantes de los países OECD que asisten a escuelas privadas, obtienen en promedio 25 puntos más en lectura que los estudiantes que asisten a escuelas públicas. Esta relación se mantiene en 15 países OECD. Sin embargo, los estudiantes que asisten a escuelas privadas también son los más favorecidos socio-económicamente, por lo que parte de la relación positiva entre escuelas privadas y rendimiento se debe a las características socio-económicas de la escuela y de los estudiantes, en lugar de ser una ventaja intrínseca de las escuelas privadas. Después de controlar la situación socio-económica y las características demográficas de los estudiantes y de las escuelas la relación se reduce a 3,4 puntos en promedio y deja de ser estadísticamente significativa. De hecho, de los 15 países OECD que muestran una relación positiva entre la asistencia a escuelas privadas y el rendimiento, sólo 3 muestran una clara ventaja por asistir a la escuela privada: Eslovenia, Canadá e Irlanda, donde los estudiantes con antecedentes similares que asisten a escuelas privadas obtienen al menos 24 puntos más en el evaluación de la lectura que los estudiantes que asisten a escuelas públicas. En cambio, en Japón y el Reino Unido, los estudiantes con antecedentes socioeconómicos similares que asisten a escuelas privadas obtienen al menos 31 puntos menos que aquellos que asisten a escuelas públicas. En Japón, una explicación común para este resultado es que algunos estudiantes que no pueden asistir a escuelas públicas reconocidas por su alto rendimiento, pueden optar por escuelas privadas como una segunda opción. (OECD, 2011b)

Tamaño de la escuela: No hay un acuerdo claro entre los investigadores y educadores sobre las condiciones para clasificar a un centro educativo como "pequeño" o "grande". Según la recopilación de investigaciones realizada por la IEA las escuelas pequeñas ofrecen un ambiente de aprendizaje más íntimo. Las escuelas pequeñas tienden a proporcionar ambientes más seguros y se caracterizan por un mejor sentido de la comunidad (Hill y Christensen, 2007; Klonsky, 2002; Wasely, Fino, Gladden, Holanda, el rey, Mosak, y Powell, 2000). Las escuelas, sin embargo, debe ser lo suficientemente grandes como para ser rentables, y prever una infraestructura de apoyo como bibliotecas, laboratorios y gimnasios, pero no tan grandes como para ser organizacionalmente complicadas para funcionar (Martin et al, 2000; Mullis et al, 2011)

Autonomía: Según PISA 2009 los países donde los centros educativos tienen mayor autonomía para definir lo que se enseña, asignar recursos y establecer el método de evaluación de los estudiantes, obtienen mejores resultados estudiantiles. También es cierto que en los países donde las escuelas rinden cuentas por medio de la publicación pública de su rendimiento, tienden a estar mejor. Los datos de PISA muestran que en los sistemas escolares donde la mayoría de los centros educativos publican sus resultados, existe una relación positiva entre la autonomía escolar (en la asignación de recursos) y el desempeño estudiantil. En un sistema escolar donde todos los datos son públicos y donde el nivel de autonomía está por encima del promedio, los estudiantes obtienen 2.6 puntos más en su capacidad de lectura que los estudiantes cuyas escuelas cuentan con un nivel medio de autonomía.

Otros estudios (Catalano, 2010), no demuestran una relación contundente entre la autonomía y el éxito académico. Una investigación que compara los resultados de la autonomía entre el sistema educativo de Inglaterra y el de Italia, describe como el sistema inglés se caracteriza por ser un sistema orientado al mercado, mientras que el italiano es más centralizado. En Inglaterra, la autonomía escolar es muy amplia, e incluye la gestión de todo el personal, mantenimiento de edificios e instalaciones, mientras que la autonomía de la escuela italiana es ciertamente más limitada. El sistema inglés parece ser eficaz, ya que las tasas de rendimiento y los retornos a la educación son más altos que los de Italia y que los del promedio de la OCDE. Sin embargo, todo esto parece estar influido por una mayor segregación social de los estudiantes en Italia, y las grandes diferencias en cuanto al pago del personal docente y el gasto dedicado a la educación (Catalano, 2010).

Inés Aguerrondo (en Blanco et al, 2008) rescata dos factores que revelan autonomía en un centro educativo: la capacidad de reclutar y despedir docentes y la autonomía en la selección del director. Un estudio analizado por esta investigadora muestra que el impacto de poder reclutar y despedir al personal docente es crucial para el éxito. El estudio demuestra, además, que es necesario un balance entre demasiada o muy poca autonomía del centro educativo. El caso de Dinamarca muestra que mucha autonomía de la escuela en la aplicación de pruebas también puede dañar el proceso de mejoramiento. Por el contrario, el caso de Grecia demuestra que la falta de autonomía de la escuela (en la toma de decisiones cotidianas, en el currículo, los tiempos, y el empleo docente) también dificulta el logro educativo. (Blanco et al, 2008).

Ambiente organizacional: El clima escolar influye sobre el rendimiento estudiantil. El clima escolar está compuesto de muchos factores, incluidos los valores, las culturas, las prácticas de seguridad y estructuras organizativas que hacen que una escuela funcione y reaccione de manera particular. El **respeto de los estudiantes y profesores, un ambiente seguro y ordenado, las interacciones constructivas entre los administradores, maestros, padres y alumnos**, todo ello contribuye a un clima escolar positivo y a generar un rendimiento más alto entre los estudiantes. Las ausencias de los docentes tienen un impacto en el rendimiento de los estudiantes pues reducen las oportunidades de los alumnos para aprender. El ambiente escolar también mejora cuando los funcionarios muestran una actitud positiva hacia los estudiantes, colaboran en las actividades curriculares y extracurriculares, y promueven el desarrollo profesional (Mullis et al, 2011).

El éxito de un centro educativo se ve influenciado por la actitud de **cooperación** entre los administradores escolares, docentes y padres de familia (en National Education Association, 2008, citado por Mullis et al, 2011). Investigaciones indican que cuando los padres participan en la educación de sus hijos, el resultado es un aumento en el rendimiento académico de los estudiantes y una mejora en la actitud general hacia la escuela (en Dearing, Kreider, y Weiss, 2008, citado por Mullis et al, 2011).

La **colaboración** entre los docentes es considerada como esencial para la creación y el mantenimiento de comunidades de aprendizaje profesional, donde se comparten ideas e innovaciones. La investigación sugiere que si los maestros trabajan más colaborativamente, el aprendizaje de los alumnos puede aumentar (en Wheelan y Kesselring, 2005, citado por Mullis et al, 2011). Los profesores que discuten su trabajo con los colegas y colaboran en la planificación y ejecución de las lecciones, por lo general se sienten menos aislados y son menos propensos a dejar la enseñanza (en Johnson, Berg, y Donaldson, 2005, citado por Mullis et al, 2011).

El perfil del docente, como trabajador individual que es capaz de resolver solo todos los problemas que enfrenta, corresponde a otra realidad de la educación. Debido a que la población estudiantil actualmente presenta una serie de características muy diversas y a la gran cantidad de estudiantes en una clase, se hace cada vez más compleja la enseñanza y generación de conocimiento y se requiere de profesores con muchas competencias personales, por lo que no siempre es posible que puedan ser desempeñadas por una sola persona. Como consecuencia, las propuestas de mejoramiento de la efectividad de los centros educativos buscan grupos de profesores que tengan la capacidad de conformar equipos docentes que compartan saberes y se complementen. Esto requiere que cada profesional de la docencia conozca sus fortalezas, pero también que pueda aceptar sus debilidades y carencias. Este nuevo modelo, requiere además hacer entender al centro educativo y a los docentes que tanto los éxitos como los problemas no son individuales, sino que pertenecen a todo el equipo. De ahí la relevancia de que el líder educativo estimule la cultura de colaboración en el trabajo de los docentes, y el compromiso para generar las condiciones reales que permitan que esto pueda materializarse. Las instituciones educativas están concebidas para el trabajo individual y esto se expresa no sólo en el edificio y la distribución de los espacios internos, sino también en las regulaciones, la

organización de los horarios y la carga de trabajo de cada profesor. Para los profesores, el trabajo en equipo implica que dentro del horario de trabajo se debe tener tiempo no sólo para estar frente a los alumnos, sino también para realizar actividades con otros docentes (Blanco et al, 2008).

Liderazgo: El resultado en el rendimiento del centro educativo también se relaciona con la forma en que el mismo es liderado. Países como Turquía, Luxemburgo, Chile, Estados Unidos, Perú, Uruguay, Argentina, Panamá y Brasil obtienen hasta 15 puntos porcentuales más que otros países en el rubro de liderazgo de las pruebas PISA. (OECD, 2010c). Además, entre los padres, profesores y políticos que trabajan para mejorar la educación, los educadores en posiciones de liderazgo son los únicos bien ubicados para asegurar la sinergia necesaria. (Seashore, Leithwood, Wahlstrom & Anderson, 2010).

El director del centro educativo desempeña un papel fundamental en el desarrollo de comunidades profesionales de aprendizaje. Las investigaciones apuntan a que el estilo de liderazgo en la escuela tiene un efecto indirecto en el rendimiento estudiantil (Bruggenkate, 2009). El tipo de liderazgo que más resultados positivos produce sobre el desempeño estudiantil es el liderazgo instruccional, pues este se encuentra más fuertemente enfocado a la efectividad estudiantil. También es importante el liderazgo que ejerce una articulación clara de la misión de la escuela y la gestión del currículo, el cual abarca diferentes dimensiones (Davies, 2009; Marzano y McNulty, 2005; Robinson, 2007). Un líder da coherencia a los propósitos del centro educativo, mediante la alineación de su estructura y su cultura con su objetivo fundamental (Dufour, Ekar, y Dufour, 2005). Esto incluye guiar al centro educativo para establecer direcciones y la búsqueda de oportunidades en el futuro, la supervisión de las metas, así como la creación de un ambiente de aprendizaje efectivo y un clima escolar positivo. (Mullis et al, 2011).

Un fuerte liderazgo educativo es particularmente relevante a la hora de lograr mejoras. Las reformas llevadas a cabo en Boston, Inglaterra y Singapur demuestran que un buen liderazgo escolar es esencial para poner en práctica cambios sustanciales y rápidos (Barber & Mourshed, 2008).

El tamaño de la clase: Muchos países han apoyado políticas para reducir la cantidad de alumnos en las aulas como apuesta para mejorar la educación. Sin embargo, las pruebas disponibles sugieren que, salvo en los primeros años escolares, la reducción de la cantidad de alumnos no tiene un fuerte impacto sobre la calidad de los resultados. De 112 estudios que analizaron el impacto de la reducción de la cantidad de alumnos por aula con relación a los resultados de los estudiantes, solo 9 hallaron una asociación positiva. Los restantes 103 no encontraron ninguna relación de importancia o incluso hallaron una relación negativa significativa. Aún en los casos donde se observó una relación significativa, su efecto no fue sustancial. (Hanushek citado en Barber & Mourshed, 2008).

En todos los países de la OCDE, es más común encontrar que escuelas con estudiantes de mejor nivel económico cuenten con clases de mayor tamaño y ofrezcan

más actividades extracurriculares. En 22 países, PISA muestra una relación positiva moderada o fuerte entre la situación socioeconómica de la escuela y el tamaño promedio de clase. En 12 países OECD, hay una relación positiva moderada o fuerte entre el desarrollo socioeconómico de la escuela y el índice de actividades extra-curriculares, que mide la cantidad de actividades extra-curriculares que se ofrecen en la escuela (OECD, 2010c). Sin embargo, algunas investigaciones indican que clases pequeñas durante los primeros años de escolaridad pueden beneficiar el desarrollo académico de los estudiantes (Mullis et al, 2011).

Entorno de aprendizaje: Los resultados de PISA en los países de la OCDE muestran que el 3% de la variación en el rendimiento escolar es explicado por diferencias en el entorno de aprendizaje, mientras que el 9% es atribuible tanto a la situación socioeconómica como al entorno de aprendizaje, ya que a mayor nivel socioeconómico, es mayor el grado de disciplina, percepción más positiva de los valores de la escuela y mayor compromiso de los docentes. Por el contrario, las escuelas cuyas poblaciones estudiantiles provienen de entornos desfavorecidos no pueden ser sometidas a la misma clase de presión de los padres para mejorar la disciplina en el salón de clases o para asegurarse de que se sustituya a profesores ausentes o desmotivados.

La influencia conjunta de la situación socioeconómica y el ambiente de aprendizaje en el rendimiento es particularmente grande en algunos países. Por ejemplo, Luxemburgo, Alemania, Japón, Turquía, Italia, la República Checa, Chile, Trinidad y Tobago, Argentina, Croacia, Uruguay, Singapur, Montenegro y Macao China, obtienen 15 o más puntos porcentuales que son explicados en forma conjunta por el ambiente de aprendizaje y la situación socioeconómica. En la mayoría de los países y economías, sólo una pequeña proporción (cinco puntos porcentuales o menos) de la variación en el rendimiento, se puede atribuir únicamente al ambiente de aprendizaje. (OECD, 2010c)

En muchos países, las escuelas con mejores climas disciplinarios, comportamientos positivos entre los docentes y mejores relaciones entre docentes y estudiantes tienden a alcanzar las puntuaciones más altas en lectura, incluso después de haber controlado la variable de situación socioeconómica. Por ejemplo, incluso controlando el contexto socio-económico y demográfico de los estudiantes y las escuelas, el desempeño de las escuelas se relaciona positivamente con el índice de las relaciones profesor-alumno en 10 países de la OECD y 7 de los países socios, también se relaciona positivamente con el índice de disciplina en 16 países OECD y 22 países asociados, y se relaciona positivamente con el índice de factores (relacionados al docente) que afectan el clima escolar en 14 países OECD y 6 de los países socios. La presión que ejercen los padres sobre los directores de los colegios para elevar los estándares académicos y el rendimiento se relaciona con un desempeño estudiantil más alto en 19 países OECD y 10 países asociados, pero después de controlar por la situación socioeconómica de los estudiantes y las escuelas se encuentra una relación positiva solo en cuatro países de la OCDE y cinco países socios.

En resumen, los estudiantes se desempeñan mejor en las escuelas que cuentan con una mayor disciplina en clase, cuyos docentes muestran un comportamiento más positivo y existen mejores relaciones profesor-alumno, en parte debido al favorable

entorno socioeconómico de los estudiantes que ayuda a reforzar un clima propicio para el aprendizaje. Las escuelas con mejores climas en cuanto a disciplina, comportamientos más positivos entre los maestros y mejores relaciones profesor-alumno tienden a lograr mejores resultados en lectura. Del total de países de la OCDE, el 81% de los estudiantes manifiestan que la mayor parte del tiempo pueden trabajar bien en clase y sólo el 23% asiste a escuelas cuyos directores informan que los estudiantes no son ayudados por los profesores (OECD, 2010c).

La International Association for the Evaluation of Educational Achievement (IEA) hace referencia a los resultados de TIMSS para ilustrar que las escuelas ubicadas en comunidades prósperas no sólo tienen un cuerpo de estudiantes más aventajados, sino que es probable que disfruten de clases pequeñas, profesores mejor capacitados y remunerados, ambientes seguros, así como del apoyo de padres bien educados y prósperos. Todos estos factores sirven para apoyar y reforzar el aprendizaje de los estudiantes, a pesar de que no sea posible separar por completo sus efectos relativos. (Martin et al, 2000).

Instalaciones y recursos: La enseñanza y el aprendizaje se facilitan cuando las instalaciones son adecuadas y se cuenta con los materiales y equipos necesarios para la consecución de objetivos de aprendizaje específicos. Resultados de las pruebas TIMSS indican que los estudiantes de las escuelas que están bien dotadas de recursos, en general, tienen un mayor rendimiento que los de las escuelas donde la escasez de recursos afecta la capacidad para aplicar el plan de estudios. Hay dos tipos de recursos, los generales y los que afectan específicamente la implementación del currículo. Los recursos generales incluyen material didáctico, el presupuesto para suministros, edificios escolares, sistemas de iluminación y espacio en las aulas. Los específicos se refieren a recursos utilizados en la enseñanza de las matemáticas y la ciencia como computadoras, software, calculadoras, equipo para laboratorio, material de biblioteca y recursos audiovisuales (Mullis et al, 2011).

Un factor fundamental asociado al progreso integral de los alumnos, especialmente en países en desarrollo, es la cantidad, calidad y adecuación de las instalaciones y recursos didácticos. Las escuelas eficaces tienen instalaciones y recursos dignos; pero, a su vez, la propia escuela los utiliza y cuida. Los datos indican que el entorno físico donde se desarrolla el proceso de enseñanza y aprendizaje tiene una importancia radical para conseguir buenos resultados. Por tal motivo es necesario que el espacio del aula esté en unas mínimas condiciones de mantenimiento y limpieza, iluminación, temperatura y ausencia de ruidos externos; también es relevante la preocupación del docente por mantener el aula cuidada y con espacios decorados para hacerla más alegre; y, como ya se ha comentado, la disponibilidad y el uso de recursos didácticos, tanto tecnológicos como tradicionales. (Blanco et al, 2008).

Aún no existe evidencia contundente sobre la eficacia de la tecnología en el salón de clases, no hay evidencia que indique que el acceso y el uso de la computadora tenga un impacto positivo en el rendimiento estudiantil (en Laffey, Espinosa, Moore, y Lodree, 2003, citado por Mullis et al, 2011). El uso eficaz de la tecnología requiere de formación profesional adecuada para profesores, estudiantes y personal de la escuela. (Mullis et

al, 2011). Sin embargo, las computadoras y el uso de Internet ofrecen a los estudiantes la oportunidad de profundizar conceptos y proporcionan el acceso a vastas fuentes de información, por lo que pueden ampliar sus oportunidades de aprendizaje.

Materiales educativos: Existe una correlación positiva entre materiales educativos y el rendimiento de los alumnos. La mayoría de estos modelos de análisis se han ejecutado después de 1980; 13 y 17 modelos presentan una relación positiva entre utilización de textos y materiales de lectura y el rendimiento de los alumnos, cuatro plantearon que no hay relación y ninguno mostró una relación negativa. Acceder a otros materiales educativos como pizarrón, globo terráqueo y mapas, también tiene un efecto positivo en el rendimiento, pero presenta una imagen menos optimista; aunque 14 de 34 modelos encontraron una relación positiva, 17 de los 34 no muestran relación alguna. La mayoría de los indicadores de infraestructura, incluyendo no sólo la calidad de la construcción sino también el mobiliario y los servicios de electricidad y agua, presentan un efecto positivo en el rendimiento (23 de 70 modelos), pero un número considerable (45 de 70) no mostraron relación alguna (Velez, Schiefelbein, Valenzuela, 1993).

Características del docente

Selección: Los sistemas educativos con más alto desempeño cuentan con mecanismos más eficientes para seleccionar a los postulantes para capacitación docente, comparados con los sistemas de bajo desempeño, y reconocen que una mala decisión en la selección puede derivar hasta en 40 años de mala enseñanza. Estos mecanismos asumen que para que una persona pueda convertirse en un docente eficiente deberá poseer cierto conjunto de características susceptibles de identificación antes de ejercer la profesión: un alto nivel general de uso de la lengua y aritmética, fuertes capacidades interpersonales y de comunicación, el deseo de aprender y la motivación para enseñar. Los procedimientos de selección de Singapur y Finlandia están entre los más eficientes. Ambos sistemas ponen fuerte énfasis en los logros académicos de los postulantes, sus habilidades de comunicación y su motivación hacia la docencia (Barber & Mourshed, 2008).

Un claro ejemplo de cómo el control de ingreso al profesorado puede tener un fuerte impacto positivo sobre la calidad de los futuros docentes puede verse en la diferencia entre el trato brindado por el sistema surcoreano a sus docentes primarios y secundarios con respecto a otros países. Para convertirse en docente primario, el postulante debe completar en primer lugar un curso de grado en educación de cuatro años en una Universidad Nacional de Educación. Las vacantes para estos cursos son limitadas, para garantizar que la oferta de docentes se ajuste a la demanda. El ingreso es por orden de mérito. La admisión a los cursos de primer grado en Corea del Sur se basa en los resultados del Examen Universitario de Ingreso Nacional; el puntaje mínimo de ingreso a los cursos de capacitación docente exige que los estudiantes se encuentren dentro del 5% superior de su promoción. Así, los cursos son altamente selectivos y sus graduados tienen muy altas probabilidades de conseguir empleo como

docentes. Esto asegura que la atracción, estatus y calidad de los cursos permanezcan altos (Barber & Mourshed, 2008).

Formación, experiencia y actitud: La evidencia disponible sugiere que el principal impulsor de las variaciones en el aprendizaje escolar es la calidad de los docentes (Barber & Mourshed, 2008). El profesor es el agente principal de la implementación del currículo y un determinante muy influyente sobre el ambiente que se desarrolla en el aula (en Lundberg y Linnakylä, 1993; Rivkin, Hanushek y Kain, 2005, citado por Mullis et al, 2011). Además, los comportamientos, actitudes, y el desempeño de los estudiantes varían de acuerdo al método de instrucción del docente (en Kurtz-Costes y Schneider, 1994, citado por Mullis et al, 2011). La investigación sugiere que, para contar con profesores de excelencia, estos deben tener conocimientos académicos elevados, enseñar en el campo en el que recibieron su formación, contar con años de experiencia, y participar en procesos de inducción de alta calidad y en programas de desarrollo profesional (en Mayer, Mullens, y Moore, 2000, citado por Mullis et al, 2011).

A partir de la revisión de 18 informes de investigación que incluyen 88 modelos de estimación o ecuaciones de regresión en países de América Latina durante los últimos 20 años, se constató en 31 de 68 modelos que el nivel educativo de los docentes está asociado significativamente con el rendimiento académico, y en 25 de 62 modelos también fue significativa la experiencia docente (Velez, Schiefelbein, Valenzuela, 1993).

De acuerdo con el estudio de McKinsey (Barber & Mourshed, 2008), los países o comunidades con alto desempeño educativo (como por ejemplo Singapur, Finlandia y Boston) tienen en común que consiguieron a las personas más aptas para ejercer la docencia (“la calidad de un sistema educativo tiene como techo la calidad de sus docentes”) y desarrollaron a estas personas hasta convertirlas en instructores eficientes (la única manera de mejorar los resultados es mejorando la instrucción). Los estudiantes de Singapur tienen el puntaje más alto en la evaluación TIMSS (Tendencias en el Estudio Internacional de Matemática y Ciencias) pese a que el gasto por estudiante primario en Singapur es inferior al de casi cualquier otro país desarrollado. En Finlandia, los alumnos no ingresan a la escuela hasta los siete años y asisten a clase solo de cuatro a cinco horas diarias durante los dos primeros años de escolaridad. No obstante, a los 15 años de edad son los mejores del mundo en pruebas de matemática, ciencias, lectura y resolución de problemas, 50 puntos por encima de sus vecinos noruegos. En Estados Unidos, Boston aumentó la cantidad de alumnos que satisfacen el estándar MCAS (Massachusetts Comprehensive Assessment System) del 25% al 74% en matemática y del 43% al 77% en inglés, en solo seis años (Barber & Mourshed, 2008).

Controlados los demás factores, la **experiencia** con la que cuenten los profesores y maestros en el ejercicio de la docencia es una de las variables que hace la diferencia, sobre todo durante los primeros años de práctica profesional (en Clotfelter, Ladd, y Vigdor, 2006; Hanushek, Kain, O'Brien y Rivkin, 2005, citado por Mullis et al, 2011).

El **desarrollo profesional** por medio de seminarios, talleres, conferencias, y publicaciones profesionales puede ayudar a los profesores a aumentar su eficacia y

ampliar sus conocimientos (en Yoon, Duncan, Lee, Scarloss, y Shapley, 2007, citado por Mullis et al, 2011). La calificación y competencias de los profesores pueden ser críticas. TIMSS ha demostrado que existe una variación considerable entre países cuyos maestros cuentan con niveles de educación completos, así como en aquellos donde los profesores de matemáticas y ciencias poseen su especialización en la materia (en Goldhaber y Brewer, 2000, citado por Mullis et al, 2011, 2011). En el siglo 21, es aún más importante que nunca que un docente, además del conocimiento pedagógico y el conocimiento de los alumnos y sus características, posea un alto conocimiento sobre las tecnologías de información (en Darling-Hammond, 2006; Ertmer, 2003; Hill & Lubienski, 2007, citado por Mullis et al, 2011).

La investigación concluye que gran parte del éxito escolar depende de las oportunidades de desarrollo profesional con las que cuentan los docentes. De tal modo que el desarrollo profesional de los docentes es de vital importancia para alcanzar mejoras dentro del sistema educativo. Si los maestros no se actualizan profesionalmente, corren el riesgo de desconocer los avances que se han producido en el tema educativo y en sus áreas disciplinares (Mullis et al, 2011).

En otras palabras, se necesita de la actualización constante, pues los sistemas educativos con más alto desempeño lo logran debido a: i. el acceso que tiene el personal docente a procesos de **capacitación**, ii. presencia de métodos efectivos de **selección** de los aspirantes más apropiados y iii. **buenos salarios** iniciales (aunque no extraordinarios). Con estas premisas se eleva el estatus de la profesión, lo que facilita la atracción de candidatos aún mejores (Barber & Mourshed, 2008).

Las **actitudes** de los profesores, tales como la motivación y la autoeficacia, contribuyen a mejorar la experiencia de aprendizaje de los estudiantes y a sus logros académicos. Los maestros que están satisfechos con su profesión y sus condiciones de trabajo en los centros educativos están más motivados para enseñar y preparar sus clases. Esto está muy relacionado con los factores mencionados anteriormente, pues los agentes desmotivadores pueden ser salarios bajos, muchas horas de enseñanza, la falta de equipo, espacio de trabajo inadecuado, la falta de comunicación y falta de colaboración entre el personal docente (Mullis et al, 2011). La investigación ha mostrado que la autoconfianza de los maestros en sus habilidades de enseñanza está asociada al rendimiento y a la motivación de los estudiantes (en Bandura, 1997; Henson, 2002, citado por Mullis et al, 2011).

Expectativa del docente sobre sus estudiantes: Según Javier Murillo, uno de los resultados más consistentes en la investigación sobre eficacia escolar, se refiere a las altas expectativas globales. Los alumnos aprenderán en la medida en que el docente confíe en que lo pueden hacer. Así, de nuevo, las altas expectativas del docente para con sus alumnos se constituyen como uno de los factores de eficacia escolar más determinantes del logro académico. Pero confiar en los alumnos no es suficiente si éstos no lo saben. De esta forma, elementos ya mencionados tales como la evaluación y, sobre todo, la comunicación frecuente de los resultados, una atención personal por parte del docente o un clima de afecto entre docente y alumnos son factores que contribuyen a que esas altas expectativas se conviertan en autoestima por parte de

estos últimos y, con ello, en altos rendimientos. Así, un profesor eficaz debe tener altas expectativas hacia sus alumnos y, además, tiene que dejar saber a sus alumnos sobre estas percepciones (Blanco et al, 2008).

Salarios: Como ya se mencionó, las escuelas efectivas requieren la combinación apropiada de personal capacitado y talentoso, planes de estudio adecuados, instalaciones en buenas condiciones y estudiantes motivados que estén dispuestos a aprender. Los resultados de PISA muestran también que salarios más altos para los profesores se asocian con un mejor rendimiento de los estudiantes, lo cual coincide con investigaciones que muestran que la calidad del profesorado mejora de forma más eficaz los resultados de los estudiantes comparada con la conformación de clases más pequeñas (OECD, 2010c).

Por tal razón, casi en todo el mundo, los principales sistemas educativos hacen dos cosas: han desarrollado mecanismos eficientes para seleccionar a los docentes que recibirán capacitación y pagan buenos salarios iniciales. Estas dos acciones tienen un claro y evidente impacto en la calidad de los futuros docentes. Estas mismas características están, por lo general, ausentes en los sistemas con bajo desempeño. Los sistemas de alto desempeño han llegado a la conclusión de que si bien aumentar los salarios en consonancia con otras actividades es importante, llevarlos por encima del promedio del mercado de graduados no conduce a aumentos futuros sustanciales en la calidad o en cantidad de aspirantes (Barber & Mourshed, 2008).

Status del docente: En todos los sistemas estudiados por McKinsey, la capacidad del sistema educativo para atraer a las personas más capacitadas a la docencia tiene estrecha relación con el estatus de la profesión. En todos los sistemas educativos existen fuertes circuitos de retroalimentación asociados al estatus de la profesión docente. Una vez que la docencia se convierte en una profesión de gran estatus, más gente talentosa se dedica a ella, llevando ese estatus a un nivel aún más alto. Esto es particularmente evidente en Finlandia y Corea del Sur, donde un cuerpo docente históricamente fuerte ha dado a la profesión gran estatus a la vista del público en general, permitiendo a estos países atraer a más postulantes de alto nivel y perpetuar ese estatus. A la inversa, allí donde la profesión tiene bajo estatus, ésta atrae aspirantes menos calificados, lo que lleva el estatus de la profesión aún más abajo y, con él, el nivel de las personas que es capaz de atraer (Barber & Mourshed, 2008).

Estudios realizados en Costa Rica

En nuestro país los estudios de factores asociados al rendimiento académico en primaria y secundaria que utilizan metodologías más rigurosas e internacionalmente aceptadas son escasos y hasta hace muy poco reflejaban el esfuerzo individual de sus autores.

Tania Moreira es una de las investigadoras que ha realizado más aportes al estudio de factores asociados al rendimiento estudiantil por medio de modelos de regresión

múltiple y multinivel. En una investigación realizada para identificar relaciones entre factores individuales e institucionales y la nota en la prueba nacional de Bachillerato de Matemática (Moreira, 2009) reporta evidencias de la asociación significativa entre esta medida de rendimiento con el historial académico y género del estudiante, estado civil de los padres, escolaridad de la madre, total de miembros familiares y espacio físico. Los resultados se refieren a una muestra probabilística, estratificada y proporcional de estudiantes provenientes de colegios públicos académicos diurnos de Costa Rica. Se utilizó un modelo de regresión múltiple.

Por su parte Ligia Rojas en su tesis doctoral (2004) explora, por medio de un modelo multinivel, factores asociados a la repitencia de estudiantes en sétimo año de colegios académicos, diurnos y públicos, con una muestra de 1442 estudiantes y 115 docentes que impartían clases de las asignaturas de Inglés, Español, Estudios Sociales, Ciencias y Matemáticas. Entre sus hallazgos se incluyen asociaciones directas con la repitencia en las siguientes variables: mala relación de los padres con el estudiante, según la percepción de este último, así como una mala interacción entre profesor y estudiante, según la valoración del profesor. Se encontró asociación inversa con las siguientes variables: autoestima, motivación, realizar mayor cantidad de tareas, recibir ayuda para realizar esas tareas, respeto percibido por los estudiantes de sus iguales, residir con ambos padres durante el tiempo lectivo, indicadores de condición socioeconómica del estudiante y su familia, e indicadores de la infraestructura e instalaciones del centro educativo, incluyendo el acceso a internet.

Finalmente, desde el año 2009 el Ministerio de Educación, específicamente el Departamento de Evaluación Académica y Certificación, de la Dirección de Gestión y Evaluación de la Calidad ha venido realizando estudios de factores asociados al rendimiento en sus pruebas de diagnóstico de sexto y noveno grados con modelos de regresión múltiple y multinivel. Se utilizan muestras estratificadas aleatorias de conglomerados completos, donde el conglomerado es el centro educativo. Se definen tres niveles de agregación: estudiante, docente y director. Los participantes de cada nivel deben completar un cuestionario de contexto para luego relacionar esas variables con el rendimiento de los estudiantes en las pruebas. El primer estudio de este tipo se realizó para identificar predictores en las pruebas diagnósticas de sexto grado aplicadas en el año 2008 (Ministerio de Educación Pública de Costa Rica, 2010).

Más recientemente el MEP ha publicado el estudio de factores asociados al rendimiento en las pruebas diagnósticas de noveno (Ministerio de Educación Pública de Costa Rica, 2012). Uno de los méritos que poseen estos estudios y que también representan una diferencia con PISA es que se conceptualizan y miden variables a nivel de aula y de profesor, pudiéndose identificar dimensiones explicativas asociadas a la mediación pedagógica y a la dinámica de la clase. Otra ventaja de estos estudios es que permiten medir factores idiosincráticos del contexto del país que pueden no ser relevantes a nivel internacional y no aparecer en los cuestionarios de PISA, por ejemplo.

Además, estas pruebas se realizan en las cinco asignaturas principales (Matemática, Español, Estudios Sociales, Ciencias e Inglés) a diferencia de las pruebas internacionales que se tienden a concentrar solo en Lenguaje, Matemática y Ciencias.

Finalmente, los constructos objeto de medición en las pruebas diagnósticas del MEP son conocimientos y destrezas específicos que se esperan como resultados directos del proceso de aprendizaje y que se derivan del currículo de cada materia, por tanto son más apropiadas como termómetro del logro de aprendizajes académicos formales establecidos en los programas de estudios, a diferencia de las pruebas PISA que son acurriculares.

Entre algunas de las variables que más consistentemente generan evidencia de asociación directa con los rendimientos en estas pruebas se pueden mencionar las siguientes: el ser alumno de un colegio privado, poseer un mayor nivel socioeconómico, nivel de atención y participación en el aula, motivación de la familia y satisfacción del director con el clima escolar. Los estudiantes varones presentan diferencias estadísticamente significativas a su favor al compararlos con las mujeres en Matemática, Estudios Sociales e Inglés. El grado académico del docente se asocia significativamente a mayores rendimientos en Matemática e Inglés. (Ministerio de Educación Pública de Costa Rica, 2012).

Específicamente como parte de este capítulo se presentan reportes más detallados para las pruebas de diagnóstico de noveno año en Matemática y Español, administradas en el 2010. No solamente se describen los desempeños obtenidos por los estudiantes, sino que también se presentan los resultados de los estudios de factores asociados al rendimiento en estas pruebas, sus autoras son respectivamente Yesenia Oviedo y Olga Muñoz.

En verdad podemos afirmar que ambas evaluaciones (pruebas nacionales de diagnóstico y PISA) se complementan la una a la otra, teniendo ambas propósitos diagnósticos y buscando generar datos para alimentar mejoras educativas, siendo PISA una prueba que se concentra en habilidades para enfrentar exitosamente las exigencias de las sociedades modernas, mientras que las pruebas de diagnóstico nacionales se enfocan en conocimientos y pueden examinar con más detalle variables específicas al contexto costarricense que pueden asociarse con los puntajes obtenidos por los estudiantes.

Representatividad de la muestra de PISA en Costa Rica

Ya hemos expresado que la mayor riqueza de PISA para alimentar discusiones productivas y contribuir a propuestas fundamentadas científicamente que mejoren nuestro sistema educativo, no está en el ranking ni en la comparación específica con cualesquiera otros países o regiones, sino en las implicancias que este diagnóstico de competencias tiene a lo interno del sistema educativo, así como la identificación de variables del estudiante y del centro educativo que al explicar o predecir los resultados, provean indicadores para el desarrollo de acciones de política educativa tendientes a mejorar esas competencias.

Sin embargo, en este apartado deseamos llamar la atención sobre un tema que amenaza, de alguna forma, tanto la validez de las comparaciones entre países como la

naturaleza del diagnóstico a lo interno del sistema educativo, y que se refiere a la población de 15 años que está representada en la muestra de PISA.

Específicamente para Costa Rica, en el año 2009, este indicador es sorprendente bajo, un 53% (Walker, 2011). Esto quiere decir que la muestra solo es representativa de un 53% de todos los jóvenes de 15 años. El nivel de exclusión (porcentaje complementario) de 47% no solo se explica a partir de la población de 15 años que no está cubierta por el sistema educativo, sino que también incluye el considerable número de jóvenes que, al tener adecuaciones curriculares, no formaron parte del marco muestral de PISA en esa oportunidad, así como los jóvenes que están en el sistema de educación abierta.

Entonces, ante la pregunta ¿cuáles son las competencias en Lectura y Alfabetización Matemática de los jóvenes costarricenses de 15 años?, los datos de PISA solo pueden responder cabalmente en torno a estimaciones que representan a los jóvenes de 15 años que están cubiertos por el sistema de educación formal, y, que no poseen adecuaciones curriculares.

Es imposible establecer cómo estos datos de desempeño variarían si se pudiera incluir ese 47% de jóvenes que no están representados en esta muestra, pero la lógica y el conocimiento de nuestro contexto nos inclina a pensar que el desempeño promedio de estos grupos en las pruebas sería menor, por haber tenido, en general, menos oportunidades de aprendizaje.

A partir de lo planteado en el informe en español de PISA 2009 (OCDE-Santillana; 2011) se puede hacer una simulación para establecer en cuanto podrían decrecer los resultados promedio de PISA si aumentara la tasa de cobertura en nuestro país. Los autores de este informe proponen la siguiente metodología:

“...incluso suponiendo que, de modo sistemático, las puntuaciones de los estudiantes excluidos hubieran sido peores que las de los que sí participaron, y que dicha relación tuviera una fuerza moderadamente alta, una tasa de exclusión del orden del 5% lo más probable es que generara una sobreestimación de las puntuaciones medias de cada país por debajo de los cinco puntos (en una escala con una media internacional de 500 puntos y una desviación típica de 100 puntos). Esta evaluación se basa en los siguientes resultados: si la correlación entre la propensión de las exclusiones y el rendimiento de los estudiantes es de 0,3, lo más probable es que las puntuaciones medias resultantes estuvieran sobreestimadas en un punto de resultado si la tasa de exclusión es del 1%, en tres puntos si la tasa de exclusión es del 5%, y en seis puntos si la tasa de exclusión es del 10%.” (p. 179).

Siguiendo entonces esta lógica hemos realizado una simulación para establecer cuáles podrían ser los promedios de Costa Rica en Lectura y Alfabetización Matemática si en vez de tener PISA 2009 una cobertura del 53% tuviéramos, alternativamente, la misma cobertura que tienen México (61%) o Chile (85%).

Los resultados de esta simulación se presentan en el siguiente cuadro. La diferencia de 32 puntos porcentuales entre las coberturas de Chile y Costa Rica, producen una

sobreestimación del valor promedio de Costa Rica igual a 19,2 unidades, pues los intervalos de 5 puntos que están contenidos en 32 son 6,4 y el puntaje promedio en la prueba baja 3 unidades por cada 5 puntos de incremento en la cobertura, así: $32/5 = 6,4$, y $6,4 \times 3 = 19,2$.

Por otra parte, la diferencia de 8 puntos porcentuales a favor de México en cobertura produce una sobrestimación del valor promedio de Costa Rica igual a 4,8 unidades = $(8/5) \times 3$.

Para realizar las pruebas de significancia estadística en donde se comparan los desempeños de estos dos países con los resultados simulados para Costa Rica, se partió de la estimación del error estándar del promedio para la muestra efectiva de Costa Rica que es de 3,2 puntos para Competencia Lectora y de 3 para Alfabetización Matemática. En el caso de la simulación con la cobertura de Chile es muy probable que si se incluyera la población excluida para aumentar la cobertura al 85% que exhibe este país, dicha muestra provocaría un aumento en la variabilidad de las puntuaciones; entonces se ha estimado, de manera liberal, que el error estándar del promedio podría subir a casi el doble del valor observado actual en ambas pruebas, es decir a un valor de 6 puntos.

Aún bajo esta estimación liberal del error estándar es claro que las diferencias entre los promedios de Chile y Costa Rica, simulando la cobertura de Chile, son estadísticamente significativas a favor del primer país. Es decir, puede afirmarse que si Costa Rica exhibiera los porcentajes de cobertura que muestra Chile los resultados promedio en ambas pruebas serían significativamente inferiores a los de Chile.

Por otra parte, en el caso de la comparación con México, como se trata de pocos puntos para la diferencia de cobertura, se considera que los valores de los errores estándar de los promedios de Costa Rica simulando la cobertura de México se mantendrían similares a los valores actuales. En este caso no hay diferencia en las pruebas de significancia estadística cuando se usan los datos observados o simulados, las conclusiones son las mismas: En Competencia Lectora el promedio de Costa Rica es significativamente superior al de México. En Alfabetización Matemática sucede exactamente lo contrario: el promedio de México es significativamente más alto que el de Costa Rica.

Cuadro 3

Resultados comparativos entre los promedios en las pruebas PISA de Competencia Lectora y de Alfabetización Matemática simulando para Costa Rica la cobertura de Chile (85%) y de México (61%)

Caso	Competencia Lectora		Alfabetización Matemática	
	Promedio	Error Estándar	Promedio	Error Estándar
Chile	449	3,1	421	3,1
Costa Rica simulado con cobertura de Chile	424	6	390	6
Costa Rica observado	443	3,2	409	3
México	425	2	419	1,8
Costa Rica simulado con cobertura de México	438	3,2	404	3

A partir de esta simulación parece evidente que la comparación directa entre los puntajes promedio brutos de Costa Rica y Chile debe hacerse con cautela, pues las diferencias de cobertura amenazan la validez de la comparación. Así, no sería correcto afirmar que virtualmente los desempeños de Costa Rica y Chile a nivel poblacional son iguales, al no existir diferencias estadísticamente significativas entre sus promedios observados.

Más bien, lo más probable es que si Costa Rica tuviera los mismos niveles de cobertura que Chile, sus desempeños promedio en las pruebas de PISA serían significativamente inferiores a los de este país.

Análisis correlacionales y de regresión multinivel con los resultados PISA 2009 en Alfabetización Matemática y Competencia Lectora: variables predictoras

En dos de los documentos que componen este capítulo Ana María Rodino y Oscar Salas, especialistas en Comprensión Lectora y Enseñanza de la Matemática, respectivamente, brindan su valoración cualitativa y análisis en torno a los relativamente pobres desempeños de nuestro país en estas dos pruebas de PISA.

Teniendo en cuenta esta situación, surgen interrogantes en torno a qué variables del estudiante y del centro educativo están asociadas a estos desempeños y si será posible emprender la tarea de generar políticas educativas que permitan incidir sobre algunas de esas variables, con el objeto de promover mayores niveles de desempeño en los dos constructos objeto de interés: competencia lectora y alfabetización matemática.

Entonces análisis correlacionales y multivariados que se presentan en esta sección cumplen dos propósitos, por un lado conocer el comportamiento de los puntajes de acuerdo con ciertas variables o categorías de población, y, por otro, tomando en cuenta

esos hallazgos, brindar pistas o sugerencias en torno a posibles medidas de política educativa que podrían implementarse con el propósito de mejorar los desempeños en ambas pruebas.

Esta sección del capítulo se divide en dos partes, en la primera se presentan y discuten resultados de análisis bivariados exploratorios. En la segunda se muestra un primer conjunto de resultados sobre variables predictoras del rendimiento en las dos pruebas, empleando modelos de regresión multinivel, y se generan algunas posibles implicaciones en términos de acciones de política educativa.

Análisis bivariados

Como primera exploración se generó la matriz de correlaciones entre los puntajes de las dos pruebas, incluyendo también las 5 subescalas que componen el puntaje total en Competencia Lectora. Los resultados mostraron que las correlaciones entre los dos constructos, Competencia Lectora y Alfabetización Matemática son mayores a 0.8, es decir, la tendencia indica que quien se desempeña bien en una de las dos áreas tiene una alta probabilidad de desempeñarse bien en la otra. Además, la escala total de Competencia Lectora arroja valores de correlación mayores a 0.9 con sus respectivas subescalas. Esto implica que todas estas escalas brindan prácticamente la misma información a nivel estadístico y por tanto, es suficiente trabajar con la escala total en los análisis de factores asociados.

Puntajes según características del colegio y sexo del estudiante

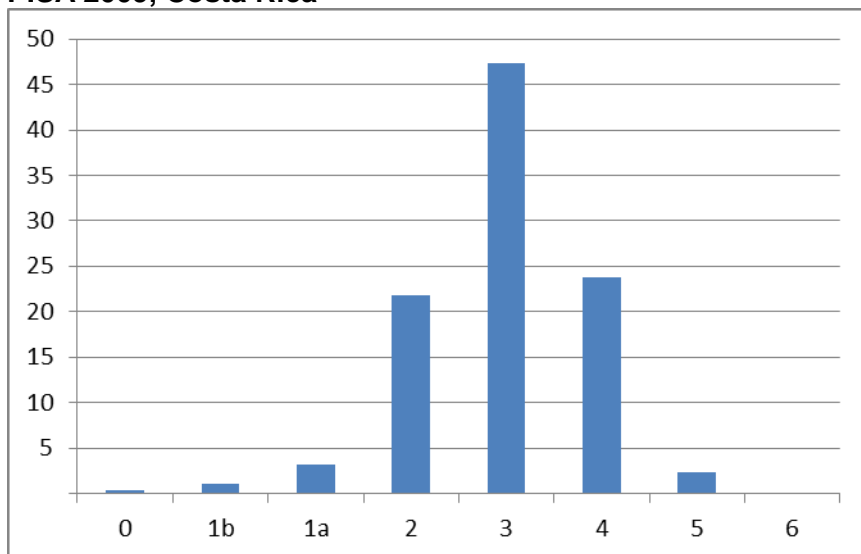
Los cuadros 4 y 5 presentan los promedios, desviaciones estándar y diferencias en los dos componentes de PISA bajo estudio, Competencia Lectora y Alfabetización Matemática, según características clasificatorias de los colegios, dependencia, zona, modalidad y horario, y según el sexo de los estudiantes examinados. Debe indicarse, además, que se realizaron pruebas de significancia estadística con el estadístico T de Student y en todos los casos se rechazó la hipótesis nula, es decir se puede afirmar que las tendencias encontradas en la muestra se mantienen a nivel poblacional.

Uno de los puntos que más llama la atención en todas estas comparaciones son los resultados por dependencia. La diferencia entre los puntajes promedio de estudiantes provenientes de colegios privados y públicos (87,72 en Lectura y 79,63 en Matemática) es, de hecho, en cada caso, mayor a la desviación estándar del conjunto total de datos. A nivel descriptivo estadístico este es un resultado contundente y muy preocupante, pues al ser tan elevadas estas discrepancias brindan evidencia para apoyar la hipótesis de que actualmente en nuestro país existen dos mundos en la educación secundaria: los centros públicos y los centros privados, que al ser tan discrepantes no representan ya una simple clasificación de la población sino, que constituyen en sí mismos poblaciones separadas y claramente diferenciadas. De manera que estas diferencias tan dramáticas en los desempeños promedio de estudiantes provenientes de colegios

públicos y privados nos entregan una nueva señal sobre lo que puede llamarse la “crisis de la educación secundaria pública”.

Para ilustrar aún más esta situación, realizamos un ejercicio generando la distribución poblacional por niveles de desempeño solo para los estudiantes de colegios privados. Con esto comprobamos que esta subpoblación en Competencia Lectora presentó la mayor frecuencia de estudiantes (casi la mitad de los examinados) en el nivel 3, mientras que en la muestra total menos de un 25% de los estudiantes se ubica en este mismo nivel, y 67,3% tienen valores iguales o inferiores al nivel 2. Por el contrario, en los estudiantes provenientes de centros privados este acumulado al nivel 2 es de solo 26,5%. Para matemática sucede algo similar, en la muestra general se acumula el 84,5% de los examinados en el nivel 2 o inferior, mientras que para los examinados de privados solo se acumula en este nivel el 51,7% de la muestra. Estos resultados se muestran en los gráficos 1 y 2.

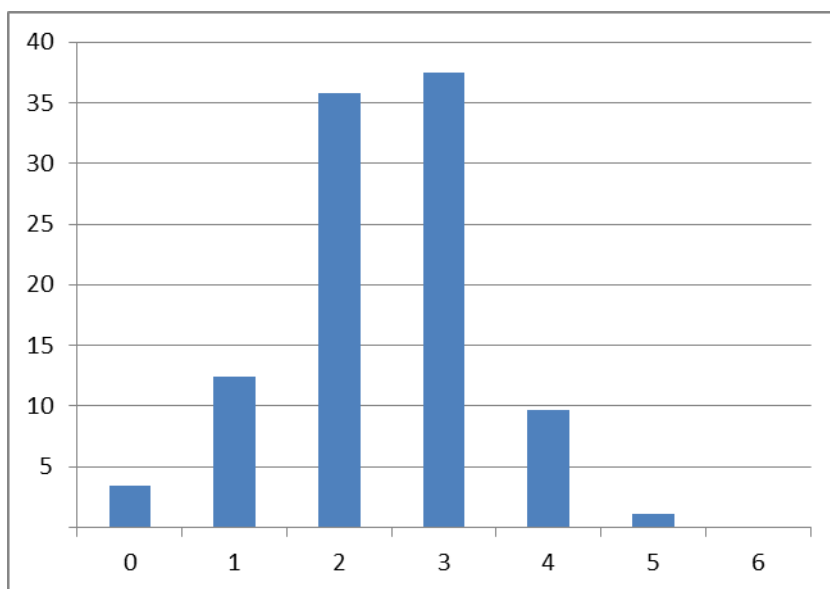
Gráfico 1
Niveles de Competencia Lectora para estudiantes provenientes de colegios privados PISA 2009, Costa Rica



Fuente: Elaboración propia

Gráfico 2

Niveles de Alfabetización Matemática para estudiantes provenientes de colegios privados PISA 2009, Costa Rica



Fuente: Elaboración propia

Sin embargo, como podrá observarse en la siguiente sección sobre los resultados del análisis multinivel de factores asociados que explican los rendimientos en estas pruebas, son otras variables y no la pertenencia a un colegio público o privado per se lo que provoca mayormente estas enormes diferencias observadas cuando se comparan ambos grupos.

Por su parte, las diferencias entre los rendimientos de estudiantes de colegios diurnos y nocturnos son también de gran magnitud y llaman a una reflexión en torno a dimensiones de excelencia y equidad asociadas a estos hallazgos, así como la realización de investigaciones para explorar sus posibles causas. Debe advertirse, eso sí, sobre la necesidad de ser cautelosos con las inferencias, pues la muestra de nocturnos solo abarcó 6 instituciones y un total de 108 estudiantes.

En el caso de Comprensión Lectora las diferencias de menor magnitud se dieron en la comparación entre hombres y mujeres, con un promedio ligeramente mayor para las segundas. En Matemática, por el contrario, el promedio fue mayor para los varones, representando esta diferencia cerca de un 40% de la desviación estándar del total del grupo, un valor bastante considerable.

En el contraste por zona los promedios fueron bastante más elevados para la región urbana, siendo esta diferencia un 56% de la desviación estándar en el caso de Competencia Lectora y un 52% en el caso de Alfabetización Matemática.

La comparación entre modalidades, aunque arrojó resultados estadísticamente significativos a favor de los colegios académicos es de relativamente poca importancia práctica en ambas pruebas.

Cuadro 4

Estadísticas Descriptivas para los puntajes de PISA 2009 en Competencia Lectora según características del colegio y sexo de los estudiantes, Costa Rica

Característica	n	Promedio	Desviación estándar
<u>Dependencia</u>			
Público	3989	432,02	68,34
Privado	589	519,74	58,26
Diferencia	3400	-87,72	10,08
<u>Zona</u>			
Urbano	3100	456,49	71,26
Rural	1478	415,65	69,64
Diferencia	1622	40,84	1,62
<u>Modalidad</u>			
Académico	3719	446,44	73,83
Técnico	859	429,73	69,21
Diferencia	2860	16,71	4,62
<u>Horario</u>			
Diurno	4470	445,09	72,31
Nocturno	108	369,21	74,62
Diferencia	4362	75,88	-2,31
<u>Sexo</u>			
Masculino	2142	436,04	76,54
Femenino	2436	449,69	69,66
Diferencia	-294	-13,65	6,88
<u>Total</u>	4578	443,3	73,27

Fuente: Elaboración propia

Cuadro 5
Estadísticas Descriptivas para los puntajes de PISA 2009 en Alfabetización Matemática según características del colegio y sexo de los estudiantes, Costa Rica

Característica	n	Promedio	Desviación estándar
<u>Dependencia</u>			
Público	3989	398,63	60,24
Privado	589	478,26	55,59
Diferencia	3400	-79,63	4,65
<u>Zona</u>			
Urbano	3100	419,76	64,56
Rural	1478	386,04	60,96
Diferencia	1622	33,72	3,6
<u>Modalidad</u>			
Académico	3719	412,04	64,89
Técnico	859	395,14	65,59
Diferencia	2860	16,9	-0,7
<u>Horario</u>			
Diurno	4470	410,31	64,78
Nocturno	108	349,5	61,39
Diferencia	4362	60,81	3,39
<u>Sexo</u>			
Masculino	2142	422,59	66,33
Femenino	2436	396,81	62,02
Diferencia	-294	25,78	4,31
<u>Total</u>	4578	408,87	65,35

Fuente: Elaboración propia

Modelo de regresión multinivel para identificar variables predictoras de los puntajes de PISA en Competencia Lectora y Alfabetización Matemática

Justificación

Específicamente en nuestro caso, para el estudio de la relación entre los puntajes de las pruebas PISA de Competencia Lectora y Alfabetización Matemática y las variables y constructos que fueron medidos por medio de los cuestionarios a estudiantes y directores, se utilizaron modelos de regresión multinivel o de niveles múltiples. Estos modelos son más poderosos que los modelos clásicos de regresión múltiple, utilizados antiguamente para analizar este tipo de relaciones.

La gran ventaja de los modelos de niveles múltiples consiste en que, a diferencia de la regresión clásica, permiten incorporar en un mismo modelo (en una misma ecuación) variables independientes pertenecientes a diferentes niveles de agregación. En nuestro caso tenemos variables individuales del estudiante (primer nivel) y variables del director y del centro educativo (segundo nivel). Como se mencionó antes, por el tipo de muestreo que utiliza PISA al interior de cada colegio (muestra simple al azar de estudiantes de 15 años) no se recolecta información del docente y sección específica donde se ubica el estudiante.

Aunque el análisis de regresión múltiple fue usado tradicionalmente para los llamados “estudios de factores asociados”, éste presenta una clara limitación cuando se trata de emplear para explicar o predecir el comportamiento humano, pues la conducta individual está influida por variables de la persona, y también por variables de su contexto. Pero los modelos clásicos de regresión no permiten de forma válida analizar conjuntamente los efectos de variables individuales y variables del contexto. (Montero, Villalobos & Valverde, 2007). Esta imposibilidad técnica tiene serias consecuencias prácticas, pues no se puede realizar una interpretación integrada de los efectos simultáneos de dichas variables y se corre el riesgo de generar conclusiones sesgadas. En este sentido hablamos de la falacia atomística cuando se incluyen solo variables del individuo y de la falacia ecológica cuando se incluyen solo variables del contexto. (Gaviria, 2005).

Por el contrario, cuando se usa un modelo de regresión multinivel (o de niveles múltiples o de efectos mixtos como se denominan actualmente en Estadística) se pueden analizar simultáneamente los efectos de las variables en los diferentes niveles de agregación. Esta característica hace que estos modelos resulten poderosos para la explicación de muchos fenómenos del comportamiento, en particular, cuando interesa identificar factores asociados al rendimiento estudiantil.

Estos modelos son relativamente nuevos. Los primeros trabajos nos refieren a Goldstein (1987) y a sus estudios de efectividad escolar. Se recomienda al lector

interesado recurrir a los siguientes textos para profundizar más en el estudio de los modelos multinivel: Raudenbush & Bryk (2002); Kreft & De Leeuw (1998) y Snijders & Bosker (1999), Gaviria (2005) y Hox (2010).

En un modelo de regresión se construye una ecuación en donde la variable dependiente o criterio (Y), aquella que interesa predecir o estimar, se concibe como una función lineal de un conjunto de variables independientes o predictoras (X1, X2, X3,...Xn). Esta función es llamada ecuación de regresión y matemáticamente se puede expresar así:

$$Y = f (X_1, X_2, X_3, \dots, X_n) = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \dots + B_nX_n + e$$

Se parte del supuesto de que el valor que observamos en la variable dependiente es el resultado de la influencia simultánea de varios factores. Las variables independientes son estos factores relevantes. Así, se busca que el conjunto de variables independientes permita predecir, con un grado razonable de precisión, el comportamiento de la variable dependiente. Esta técnica de la estadística multivariable permite también analizar la fuerza o intensidad de la asociación entre las variables. Estima el grado de relación entre cada una de las variables independientes y la dependiente, “controlando”, “aislando” o “o manteniendo constante” el efecto de las otras variables independientes en el modelo. Una vez que el modelo es ajustado, es decir, una vez que se ha estimado la ecuación de regresión que mejor reproduce el comportamiento de los datos empíricos, se puede juzgar la importancia de cada variable independiente en la predicción de la dependiente. Esto se hace examinando la magnitud de los llamados "coeficientes de regresión". (Montero, 2001). En esta ecuación “e” es el término de error, porque evidentemente es imposible predecir de manera exacta el valor de Y a partir del conjunto de variables independientes. De hecho este es el término que se intenta minimizar como parte del proceso de estimación del modelo de regresión.

Los valores B1, B2, B3... Bn son los coeficientes de regresión y representan los pesos o ponderaciones de las correspondientes variables independientes X1,X2,X3,...,Xn. El criterio fundamental que guía el procedimiento de estimación de estos valores en una ecuación de regresión específica es la minimización de los residuos o errores, definiendo éstos como la diferencia entre los valores observados de la variable dependiente y los valores estimados con base en el modelo de regresión.

La interpretación sustantiva de Bi cambia dependiendo de si la variable independiente respectiva es métrica o categórica. Si la variable independiente es métrica (se mide en unidades, como el grado o nivel que cursa el estudiante o el puntaje en una escala de lectura) y Bi es positivo, existe una relación directa entre Xi y Y (cuando aumenta una, aumenta la otra), manteniendo constantes o controlando el efecto de las otras variables independiente en el modelo. Por el contrario, si el coeficiente de regresión Bi es negativo y la variable independiente es métrica, la relación entre Xi y Y es inversa, es decir al aumentar una disminuye la otra. O sea, en los casos de variable independiente métrica estos coeficientes también se interpretan en el mismo sentido que una

pendiente en una función lineal, cada uno de ellos indica en cuántas unidades cambia Y por unidad de cambio en X_i . (Montero, 2001).

Si, por el contrario, la X_i representa una variable categórica (como el sexo o el colegio de procedencia del estudiante) se debe recurrir a la creación de variables dicotómicas o dummy con codificación 0-1, para introducirlas al modelo de regresión, donde 0 representa la categoría de referencia (definida por el investigador) y 1 la categoría de interés (igualmente definida por el investigador), de esta forma los coeficientes B_i estimados por el modelo de regresión para la variable independiente dicotómica se interpretan como la diferencia en los promedios de la Y , la variable dependiente, entre aquellos que se ubican en la categoría de interés versus los que están ubicados en la categoría de referencia. Más abajo se ilustra esta interpretación para el modelo multinivel que intenta predecir el puntaje de PISA en Competencia Lectora.

La medida R^2 , por su parte, se denomina coeficiente de determinación y se emplea para evaluar sustantivamente la capacidad predictiva del modelo de regresión. El valor de R^2 indica el porcentaje de la variabilidad en la variable dependiente que puede ser explicada a partir de su relación lineal con las variables independientes. Cuando se realizan análisis multinivel en investigación educativa es importante establecer cuál es el peso o ponderación en el R^2 de las variables individuales del estudiante versus las variables del docente y del centro educativo, pues se puede argumentar que en aquellos casos donde son más importantes las variables del contexto, comparadas con las variables individuales del estudiante, es posible intervenir de manera más directa para promover mejoras en el rendimiento.

Los coeficientes “ B ” no pueden compararse directamente, pues en general no están en las mismas unidades de medida, entonces para comparar la importancia relativa en la predicción de las variables independientes se deben analizar los valores de los coeficientes de regresión estandarizados (o coeficientes *beta*) para cada una de ellas. Estos coeficientes pueden compararse directamente de manera válida, ya que la estandarización ha eliminado el efecto de las diferentes unidades de medida de las variables. El coeficiente *beta* está entre -1 y 1 , y nos indica en cuántas desviaciones estándar cambia la variable dependiente cuando la variable independiente en cuestión cambia en una desviación estándar. Aquellas variables con un valor de *beta* más alto en valor absoluto tienen mayor importancia, más peso, en la predicción de la variable dependiente.

Debe indicarse que la interpretación sustantiva de los coeficientes de regresión y pruebas de significancia es la misma para los modelos clásicos y multinivel. Lo que cambia son las estimaciones numéricas de los errores estándar asociados a los coeficientes de la regresión, así como la probabilidad asociada de las pruebas de inferencia estadística. Normalmente, la estructura multinivel provoca que las estimaciones tengan menor precisión comparada con las se tendría si se estuviera trabajando con una muestra simple al azar. Por tanto, en el caso multinivel, las probabilidades asociadas a los coeficientes en las pruebas de inferencia estadística son más altas, teniéndose menos posibilidad de rechazar la hipótesis nula. En otras

palabras, si se ignora la estructura multinivel se concluirá erróneamente que algunos resultados son estadísticamente significativos cuando en realidad no lo son.

Aunque en un análisis de regresión multinivel los datos pueden interpretarse solamente en términos correlacionales y no causales, es importante mencionar que si el referente teórico está dado en términos de un modelo causal, este tipo de análisis ayuda a establecer si hay evidencias empíricas para apoyar la hipótesis de causalidad en una de sus condiciones necesarias, la existencia de asociación.

Metodología

En este estudio se definieron los puntajes en las pruebas de Competencia Lectora y Alfabetización Matemática como las variables dependientes, y como variables predictoras o explicativas, factores individuales del estudiante en el primer nivel y características del centro educativo y del director en el segundo nivel.

La investigación puede clasificarse como observacional transversal por su diseño y como exploratoria-correlacional por su propósito sustantivo.

Se tomaron las bases de datos globales públicas de las pruebas PISA 2009 disponibles para descargar en el sitio web de esta organización. De éstas se seleccionaron los registros correspondientes a Costa Rica. Luego se concatenaron las bases de datos de los cuestionarios de estudiantes y directores con la base de resultados de las pruebas.

De los cuestionarios se seleccionaron aquellas variables que, de acuerdo al criterio del equipo investigador, eran relevantes y aplicables para el contexto costarricense, tomando en cuenta aspectos de teoría y de la revisión de literatura que se presentó antes, así como también las condiciones específicas de nuestro sistema educativo.

Seguidamente se procedió a realizar las transformaciones y recodificaciones necesarias de los datos para que pudieran introducirse válidamente en el modelo multinivel. Como paso previo al ajuste del modelo se eliminaron del análisis aquellas variables con un porcentaje relativamente alto de valores faltantes o aquellas que presentaron muy escasa variabilidad.

Los análisis de datos se realizaron con los paquetes SPSS y STATA. En este último se utilizó un algoritmo especializado para realizar regresión multinivel con los datos de PISA.

En el cuadro 6 se presentan y definen los indicadores utilizados en el análisis. Conocer estas definiciones operativas permite al lector informado interpretar de manera directa los resultados de los modelos a partir de lo que muestran los cuadros 7 y 8 que se presentan seguidamente. Al final de la definición de cada variable se indica si ésta es dummy, el resto de las variables se tratan como métricas en el modelo.

Definiciones operativas de los indicadores utilizados en los modelos de regresión multinivel para predecir los puntajes en PISA 2009 en Competencia Lectora y Alfabetización Matemática

Observación: Si se trata de una variable dummy (con codificación 0-1, que se usa para introducir en el modelo multinivel a variables categóricas o nominales) se indica explícitamente al final de la definición de cada variable. El resto de las variables se tratan como métricas en el modelo.

Variables independientes:

Nivel: Estudiante

- **Grado:** Grado que cursa el (la) estudiante, su rango va de 7 a 12.
- **Ciclo:** Ciclo que cursa el (la) estudiante. Los valores que toma la variable son:
 - 1=Educación diversificada o técnica media
 - 0= III Ciclo
- **Sexo:** Sexo del (la) estudiante. Los valores de la variable son:
 - 1=Mujer
 - 0=Hombre
- **Asistió a preescolar:** El (la) estudiante asistió a preescolar. Los valores de la variable son:
 - 1=Sí
 - 0=No
- **Edad en primer grado:** Edad del (la) estudiante en primer grado, su rango va de 4 a 10 años.
- **Frecuencia de repetición:** Índice de frecuencia de repetición, su rango va de 1 a 3. El índice se obtuvo a partir del promedio de los siguientes ítems:
 - En educación primaria
 - En III Ciclo de la Educación General Básica
 - En educación media

El rango va de 1 a 3 porque dichos ítems pueden tomar los siguientes valores:

1= Nunca

2=Sí, una vez

3=Sí, dos veces o más

- **Padre y madre viven con estudiante:** Los valores de la variable son:
 - 1= Si padre y madre viven con estudiante
 - 0=Otro caso
- **Nivel educativo de la madre:** Nivel educativo de la madre, los valores de la variable son:
 - 0= No completó primaria
 - 1= Primaria
 - 2= III Ciclo

- 3= Educación diversificada
 - 4= Estudios generales y técnicos
 - 5= Bachillerato o Licenciatura
 - 6= Maestría o Doctorado
- **Clasificación de ocupación del padre:** Clasificación de ocupación del padre, los valores de la variable son:
 - 1= Ocupación manual de baja calificación
 - 2= Ocupación manual de alta calificación
 - 3= Ocupación no manual de baja calificación
 - 4= Ocupación no manual de alta calificación.
- **Nivel educativo del padre:** Nivel educativo del padre, los valores de la variable son:
 - 0= No completó primaria
 - 1= Primaria
 - 2= III Ciclo
 - 3= Educación diversificada
 - 4= Estudios generales y técnicos
 - 5= Bachillerato o Licenciatura
 - 6= Maestría o Doctorado
- **Padres extranjeros:** Los valores de la variable son:
 - 1= Al menos uno de los padres es extranjero
 - 0= Ambos padres son costarricenses
- **Tenencia de bienes:** Puntaje del (la) estudiante en una escala de tenencia de bienes, su rango va de 0 a 1, y se compone de 14 ítems. El valor de la medida de confiabilidad, Alfa de Cronbach, es de 0.811. Los bienes que conforman la escala son:
 - Escritorio para estudiar
 - Un lugar tranquilo para estudiar
 - Una computadora que puedes usar para hacer los trabajos del colegio
 - Programas educativos parra computadora
 - Conexión a internet
 - Libros de literatura clásica
 - Libros de poesía
 - Obras de arte
 - Libros de ayuda para tu trabajo escolar
 - Manuales técnicos
 - DVD o VHS
 - Cámara digital
 - Antena parabólica o televisión por cable
 - Enciclopedia

El rango va de 0 a 1 porque cada ítem puede tomar los valores:

0=No

1=Sí

Tenencia de elementos: Índice de tenencia de elementos en casa, su rango va de 0 a 3. El índice se obtuvo a partir del promedio de los siguientes ítems:

- Teléfono celular
- Televisor
- Computadora
- Carro
- Baño con ducha o tina

Dichos ítems podía tomar los siguientes valores:

0= Ninguno

1=Uno

2=Dos

3=Tres o más

- **Cantidad de libros en la casa:** Cantidad de libros en la casa, los valores de la variable son:
 - 1= 0 a 10 libros
 - 2= 11 a 25 libros
 - 3= 26 a 100 libros
 - 4= 101 a 200 libros
 - 5= 201 a 500 libros
 - 6= Más de 500 libros
- **Tiempo de lectura por placer:** Tiempo de lectura por placer, los valores de la variable son:
 - 1= Yo no leo por placer
 - 2= 30 minutos o menos por día
 - 3= Más de 30 pero menos de 60 minutos por día
 - 4= De 1 a 2 horas por día
 - 5= Más de 2 horas por día
- **Actitud hacia la lectura:** Puntaje del (la) estudiante en una escala de actitud hacia la lectura, su rango va de 1 a 4, y se compone de 9 ítems. El valor de la medida de confiabilidad, Alfa de Cronbach, es de 0.862. Los ítems que conforman la escala son:
 - Solamente leo si lo tengo que hacer (se recodificó para cambiar su dirección)
 - La lectura es uno de mis pasatiempos preferidos
 - A mí me gusta hablar con la gente sobre libros
 - Me gusta que me regalen libros
 - Para mí la lectura es una pérdida de tiempo (se recodificó para cambiar su dirección)
 - Me gusta ir a una librería o a una biblioteca
 - Yo sólo leo para buscar la información que necesito
 - Me gusta dar mi opinión sobre los libros que he leído
 - Me gusta intercambiar libros con mis amigos

El rango va de 1 a 4 porque cada ítem puede tomar los valores:

- 1=Completamente en desacuerdo
- 2=En desacuerdo
- 3=De acuerdo
- 4=Completamente de acuerdo

- **Frecuencia de lectura por iniciativa propia:** Índice de frecuencia de lectura de distintos tipos de texto por iniciativa propia, su rango va de 1 a 5.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- Revistas
- Tiras cómicas
- Libros de ficción (novelas, narraciones, relatos)
- Libros que no sean de ficción (didácticos, documentales, informativos)
- Periódicos

El rango va de 1 a 5 porque los ítems pueden tomar los valores:

- 1=Nunca o casi nunca
- 2=Algunas veces por año
- 3=Alrededor de una vez por mes
- 4=Varias veces por mes
- 5=Varias veces por semana

- **Frecuencia de lectura en línea:** Índice de frecuencia de lectura en línea, su rango va de 0 a 4. El índice se obtuvo a partir del promedio de los siguientes ítems:

- Leer el correo electrónico
- Chatear en línea
- Leer noticias en línea
- Utilizar un diccionario o una enciclopedia en línea
- Buscar información en línea sobre un tema en particular
- Participar en debates y foros en línea
- Buscar información práctica en línea

El rango va de 1 a 5 porque los ítems pueden tomar los valores:

- 0=No sé de qué se trata
- 1=Nunca o casi nunca
- 2=Varias veces por mes
- 3=Varias veces por semana
- 4=Varias veces por día

- **Uso de técnicas "memorísticas" al estudiar:** Puntaje del (la) estudiante en una escala de uso de técnicas "memorísticas" al estudiar, su rango va de 1 a 4, y se compone de 7 ítems. El valor de la medida de confiabilidad, Alfa de Cronbach, es de 0.754. Los ítems de la escala son:

- Cuando estudio, trato de memorizar todos los puntos tratados en el texto
- Cuando estudio, empiezo por determinar exactamente aquello que necesito aprender
- Cuando estudio trato de memorizar el mayor número posible de detalles
- Cuando estudio, vuelvo a leer el texto tantas veces que logro recitarlo de memoria

- Cuando estudio, compruebo si comprendo lo que he leído
- Cuando estudio, vuelvo a leer el texto una y otra vez
- Cuando estudio, compruebo si he retenido los puntos más importantes del texto

El rango va de 1 a 4 porque los ítems toman los valores:

- 1=Casi nunca
- 2=A veces
- 3=A menudo
- 4=Casi siempre

- **Uso de técnicas "analíticas" al estudiar:** Puntaje del (la) estudiante en una escala de uso de técnicas "analíticas" al estudiar, su rango va de 1 a 4, y se compone de 6 ítems. El valor de la medida de confiabilidad, Alfa de Cronbach, es de 0.767. Los ítems de la escala son:
 - Cuando estudio, trato de relacionar la nueva información con aquello que ya he aprendido en otras materias
 - Cuando estudio, trato de ver de qué manera esta información podría serme útil fuera del colegio
 - Cuando estudio, trato de descubrir aquellas nociones que aún no he comprendido muy bien
 - Cuando estudio, trato de comprender mejor el contenido relacionándolo con mi experiencia personal
 - Cuando estudio, trato de ver si la información del texto corresponde a aquello que sucede en la realidad
 - Cuando estudio y no comprendo algo, busco información para tratar de aclarar el punto

El rango va de 1 a 4 porque los ítems toman los valores:

- 1=Casi nunca
- 2=A veces
- 3=A menudo
- 4=Casi siempre

- **Número de lecciones de Español:** Número de lecciones de Español a la semana, su rango va de 0 a 10.
- **Número de lecciones de Matemática:** Número de lecciones de Matemática a la semana, su rango va de 0 a 10.
- **Lecciones fuera de horario escolar:** Índice de asistencia a lecciones fuera de horario escolar, su rango va de 0 a 1.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- Lecciones de profundización en Español
- Lecciones de profundización en Matemáticas
- Lecciones de profundización en Ciencias
- Lecciones de profundización en otras materias escolares
- Lecciones de recuperación en Español
- Lecciones de recuperación en Matemáticas
- Lecciones de recuperación en Ciencias
- Lecciones de recuperación en otras materias

- Lecciones para mejorar tus métodos de estudio

El rango va de 0 a 1 porque cada ítem puede tomar los valores:

0=No

1=Sí

- **Percepción sobre el valor del colegio:** Índice de percepción sobre el valor del colegio para el (la) estudiante, su rango va de 1 a 4.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- El colegio no ha hecho gran cosa para prepararme para la vida adulta cuando termine la escuela (se recodificó para cambiar su dirección)
- El colegio ha sido una pérdida de tiempo (se recodificó para cambiar su dirección)
- El colegio ha contribuido en darme confianza para tomar decisiones
- El colegio me ha enseñado cosas que podrían serme útiles para trabajar

El rango va de 1 a 4 porque cada ítem puede tomar los valores:

1=En total desacuerdo

2=En desacuerdo

3=De acuerdo

4=Totalmente de acuerdo

- **Percepción acerca del trato que le dan los profesores:** Índice de percepción del (la) estudiante acerca del trato de los profesores hacia él o ella, su rango va de 1 a 4.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- Yo me llevo bien con la mayoría de mis profesores
- La mayoría de mis profesores están interesados en mi bienestar
- La mayoría de mis profesores escucha realmente lo que yo quiero decir
- Si necesito una ayuda extra, la recibiré de mis profesores
- La mayoría de mis profesores me trata correctamente

El rango va de 1 a 4 porque cada ítem puede tomar los valores:

1=En total desacuerdo

2=En desacuerdo

3=De acuerdo

4=Totalmente de acuerdo

- **Número de estudiantes en la clase de Español:** Número de estudiantes en clase de Español, su rango va de 1 a 50.

- **Ambiente negativo en clase:** Índice de ambiente en clase de Español, valores altos implican un ambiente negativo en la clase de Español, su rango va de 1 a 4.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- Los estudiantes no escuchan lo que dice el profesor
- Hay ruido y desorden
- El profesor debe esperar largo rato para que los estudiantes se calmen
- Los estudiantes no pueden trabajar bien
- Los estudiantes empiezan a trabajar mucho después del comienzo de la clase

El rango va de 1 a 4 porque cada ítem puede tomar los valores:

- 1=Nunca o casi nunca
- 2=En algunas clases
- 3=En la mayoría de las clases
- 4=En todas las clases

- **Percepción sobre técnicas pedagógicas para el estudio de textos:** Índice de percepción sobre técnicas pedagógicas aplicadas por el docente en la clase de español para el estudio de textos, su rango va de 1 a 4.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- El profesor les pide a los estudiantes que expliquen el sentido de un texto
- El profesor hace preguntas que retan a los estudiantes a reflexionar y a comprender mejor un texto
- El profesor da a los estudiantes suficiente tiempo para reflexionar sobre sus respuestas
- El profesor les recomienda a los estudiantes leer un determinado libro o autor
- El profesor estimula a los estudiantes para que expresen su opinión sobre un texto
- El profesor ayuda a los estudiantes a ver la relación entre los relatos que leen y su propia vida
- El profesor les enseña a los estudiantes cómo relacionar la información de un texto con sus vivencias

El rango va de 1 a 4 porque cada ítem puede tomar los valores:

- 1=Nunca o casi nunca
- 2=En algunas clases
- 3=En la mayoría de las clases
- 4=En todas las clases

- **Percepción sobre técnicas pedagógicas en general:** Índice de percepción sobre técnicas pedagógicas generales aplicadas por el docente en la clase de Español, su rango va de 1 a 4.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- El profesor explica a los estudiantes lo que él espera de ellos
- El profesor se asegura de que los estudiantes estén concentrados cuando realizan una actividad de lectura
- El profesor comenta el trabajo de los estudiantes una vez terminada su actividad de lectura
- El profesor explica previamente a los estudiantes cómo va a ser evaluado su trabajo
- El profesor pregunta si todos los estudiantes entendieron cómo realizar la actividad de lectura
- El profesor califica el trabajo de los estudiantes
- El profesor les da a los estudiantes la posibilidad de hacer preguntas sobre la actividad de lectura
- El profesor hace preguntas que producen en los estudiantes ganas de participar activamente

- Apenas terminan los estudiantes su actividad de lectura, el profesor les dice en seguida si está bien o no

El rango va de 1 a 4 porque cada ítem puede tomar los valores:

- 1=Nunca o casi nunca
- 2=En algunas clases
- 3=En la mayoría de las clases
- 4=En todas las clases

- **Actividades que realiza cuando va a la biblioteca:** Índice de frecuencia de actividades que realiza el (la) estudiante cuando va a la biblioteca, su rango va de 0 a 4.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- Sacar libros para leer con el fin de distraerse
- Sacar libros para el trabajo escolar
- Hacer tareas, realizar un trabajo o buscar documentación para las clases
- Leer revistas o periódicos
- Leer libros por diversión
- Informarse sobre temas sin ninguna relación con las clases
- Navegar en internet

El rango va de 0 a 4 porque cada ítem puede tomar los valores:

- 0=Nunca
- 1=Algunas veces al año
- 2=Alrededor de una vez por mes
- 3=Varias veces por mes
- 4=Varias veces por semana

- **Hay biblioteca en el colegio:** Hay biblioteca en el colegio. Los valores que toma la variable son:

- 0= No hay
- 1= Sí hay

- **Percepción de eficacia de estrategias para comprender texto:** Índice de percepción de la eficacia de distintas estrategias para comprender un texto, su rango va de 1 a 6.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- Me concentro en las partes del texto que son fáciles de entender (se recodificó para cambiar su dirección)
- Leo rápidamente todo el texto dos veces (se recodificó para cambiar su dirección)
- Después de haber leído el texto, discuto sobre su contenido con otras personas
- Subrayo los pasajes más importantes del texto
- Resumo el texto con mis propias palabras
- Le leo el texto en voz alta a otra persona

El rango va de 1 a 6 porque cada ítem puede tomar valores entre 1 y 6, donde 1 significa que es una estrategia totalmente ineficaz y 6 significa que es muy eficaz.

- **Percepción de eficacia de estrategias para resumir un texto:** Índice de percepción de la eficacia de distintas estrategias para resumir un texto, su rango va de 1 a 6.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- Escribo un resumen. Luego, verifico que cada párrafo del texto esté mencionado en mi resumen porque es necesario que el contenido de cada párrafo esté presente (se recodificó para cambiar su dirección)
- Trato de copiar al pie de la letra la mayor cantidad de frases posible (se recodificó para cambiar su dirección)
- Antes de escribir el resumen, leo el texto tantas veces como sea posible (se recodificó para cambiar su dirección)
- Verifico cuidadosamente si los elementos más importantes del texto aparecen en mi resumen
- Leo el texto subrayando las frases más importantes. Luego, las reescribo con mis propias palabras como resumen

El rango va de 1 a 6 porque cada ítem puede tomar valores entre 1 y 6, donde 1 significa que es una estrategia totalmente ineficaz y 6 significa que es muy eficaz.

- **Frecuencia de lectura de textos para el colegio:** Índice de frecuencia de lectura de distintos tipos de textos para el colegio, su rango va de 1 a 4.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- Textos de información sobre escritores o libros
- Poesía
- Textos que incluyen diagramas o mapas
- Ficción
- Reportes de periódicos y artículos de revistas
- Instrucciones o manuales que indican cómo construir o hacer algo
- Textos que incluyen tablas o gráficos
- Material publicitario

El rango va de 1 a 4 porque cada ítem puede tomar los valores:

1=Nunca

2=Una vez

3=Dos o tres veces

4=Varias veces

- **Frecuencia de realización de tareas de lectura para el colegio:** Índice de frecuencia de realización de distintos tipos de tareas de lectura para el colegio, su rango va de 1 a 4.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- Encontrar información en un gráfico, diagrama o tabla
- Explicar la causa de los eventos de un texto
- Explicar la forma en que los personajes se comportan en un texto
- Aprender acerca de la vida del autor
- Explicar el propósito de un texto
- Memorizar un texto
- Aprender acerca del lugar de un texto en la historia de la literatura

- Describir la forma en que la información de una tabla o un gráfico está organizada
- Explicar la conexión entre las diferentes partes de un texto

El rango va de 1 a 4 porque cada ítem puede tomar los valores:

1=Nunca

2=Una vez

3=Dos o tres veces

4=Varias veces

Nivel: Institución

- **Comunidad donde está la institución:** Tipo de comunidad donde está la institución. Los valores que toma la variable son:
 - 1= Área rural (menos de 3.000 personas)
 - 2= Población pequeña (3.000 a 15.000 personas)
 - 3= Ciudad pequeña (15.000 a 100.000 personas)
 - 4= Ciudad mediana (100.000 a 1.000.000 personas)
 - 5= Ciudad grande (con más de 1.000.000 personas)
- **Instituciones compitiendo por estudiantes:** Hay otras instituciones compitiendo por los estudiantes, los valores que toma la variable son:
 - 1= Sí hay al menos una
 - 0= No hay otra institución compitiendo.
- **Total estudiantes:** Total estudiantes en la institución al 1° mayo del 2010, su rango va de 39 a 2839.
- **Porcentaje de estudiantes que repitió de 7 a 9:** Porcentaje de estudiantes de 7° a 9° que repitió el pasado año lectivo.
- **Porcentaje de estudiantes que repitió de 10 a 12:** Porcentaje de estudiantes de 10° a 12° que repitió el pasado año lectivo.
- **Razón de computadoras por estudiante:** Razón de computadoras por estudiante, su rango va de 0 a 1.88.

Dificultades para desarrollar enseñanza: Índice de factores asociados a dificultades para desarrollar la enseñanza, su rango va de 1 a 4. Valores altos en el índice indican que hay más problemas que afectan la capacidad para impartir enseñanza.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- Falta de profesores calificados de Ciencias
- Falta de profesores calificados de Matemática
- Falta de profesores calificados de Español
- Falta de profesores calificados de otras materias
- Falta de personal de biblioteca
- Falta de personal auxiliar
- Escasez o inadecuación de los laboratorios de ciencias
- Escasez o inadecuación del material pedagógico

- Escasez o inadecuación de computadoras para el trabajo en clase
- Escasez o insuficiencia de la conexión a Internet
- Escasez o inadecuación del software para trabajo en clase
- Escasez o inadecuación de los recursos de la biblioteca
- Escasez o inadecuación de equipos audiovisuales

El rango va de 1 a 4 porque cada ítem puede tomar los valores:

- 1=De ninguna manera
- 2=Muy poco
- 3=En cierta medida
- 4=Mucho

- **Actividades extracurriculares disponibles para estudiantes de 9°:** Índice de actividades extracurriculares disponibles en la institución para estudiantes de 9° año, su rango va de 0 a 1.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- Banda, orquesta o coral
- Obra de teatro o comedia musical
- Redacción del anuario, periódico o revista de la institución
- Voluntariado o actividades humanitarias, por ejemplo servicio social / brigadas de salud
- Club de lectura
- Club de debates o torneos de oratoria
- Club de estudiantes o campeonato escolar organizado alrededor de una lengua extranjera, de Matemáticas o de Ciencias
- Club académico
- Club artístico o actividades artísticas
- Equipo deportivo o actividades deportivas
- Conferencias y/o seminarios (en donde el invitado sea, por ejemplo, un escritor o un periodista)
- Colaboración con las bibliotecas locales
- Colaboración con los periódicos locales
- Giras educativas

El rango va de 0 a 1 porque cada ítem puede tomar los valores:

- 0=No
- 1=Sí

- **Clima escolar negativo:** Escala de clima escolar negativo, su rango va de 1 a 4, y se compone de 12 ítems. Valores altos representan un clima más desfavorable. El valor de la medida de confiabilidad, Alfa de Cronbach, es de 0.914. Los ítems que conforman la escala son:

- Bajas expectativas de los profesores hacia los estudiantes
- El ausentismo de los estudiantes
- Malas relaciones entre estudiantes y profesores
- Estudiantes que interrumpen las clases
- Profesores que no identifican las necesidades individuales de los estudiantes
- Al ausentismo de los profesores
- Estudiantes que se ausentan en las clases sin justificación

- La falta de respeto de los estudiantes hacia los profesores
- Personal que se resiste a los cambios
- El consumo de alcohol o de sustancias ilegales por los estudiantes
- Profesores que son demasiado estrictos con los estudiantes
- Estudiantes que amenazan o intimidan a otros estudiantes

El rango va de 1 a 4 porque cada ítem puede tomar los valores:

1=De ninguna manera

2=Muy poco

3=En cierta medida

4=Mucho

- **Expectativas de los padres acerca de la institución:** Expectativas de los padres acerca de la institución, los valores que toma la variable son:
 - 1= La institución no está sometida a una verdadera presión de los padres
 - 2= La presión proviene de una minoría de padres
 - 3= La institución es sometida a una presión constante por parte de muchos padres.

- **Uso de datos de rendimiento académico en rendición de cuentas:** Índice de uso de datos de rendimiento académico en distintos procesos de rendición de cuentas, su rango va de 0 a 1.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- La información relativa a rendimiento académico se da públicamente
- La información relativa a rendimiento académico se utiliza para evaluar el desempeño de los profesores
- La información relativa a rendimiento académico se utiliza para evaluar el desempeño de los profesores
- La información relativa a los resultados escolares es utilizada para la toma de decisiones concerniente a la asignación de recursos pedagógicos para la institución
- La información relativa a los resultados escolares es objeto de seguimiento en el tiempo por un ente administrativo

El rango va de 0 a 1 porque cada ítem puede tomar los valores:

0=No

1=Sí

Uso de métodos para monitorear la práctica de profesores de Español: Índice de uso de distintos métodos para monitorear la práctica docente de los profesores de Español, su rango va de 0 a 1.

El índice se obtuvo a partir del promedio de los siguientes ítems:

- Pruebas o controles que evalúan los desempeños de los estudiantes
- Evaluación mutua entre colegas
- Observación de las clases por el director de la institución o por profesores con experiencia
- Observación de las clases por asesores supervisores o por otras personas externas de la institución

El rango va de 0 a 1 porque cada ítem puede tomar los valores:

0=No

1=Sí

- **Escala de gestión del director:** Escala de gestión del director en la institución, su rango va de 1 a 4, y se compone de 13 ítems. El valor de la medida de confiabilidad, Alfa de Cronbach, es de 0.827.

Los ítems que conforman la escala son:

- Trato que las actividades de desarrollo profesional de los profesores estén de acuerdo con los objetivos de enseñanza de la institución
- Velo por que los profesores trabajen de acuerdo con los objetivos pedagógicos de la institución
- Voy a los salones para observar las clases
- Me sirvo de los resultados de los estudiantes para elaborar los objetivos pedagógicos de la institución
- Doy consejo a los profesores sobre la manera como pueden mejorar sus prácticas
- Sigo de cerca el trabajo de los estudiantes
- Cuando un profesor tiene un problema en clase, tomo la iniciativa de discutir de ello con él / ella
- Informo a los profesores sobre las posibilidades de actualizar sus conocimientos y competencias
- Me cerciero de que las actividades llevadas a cabo en clase estén de acuerdo con nuestros objetivos pedagógicos
- Tengo en cuenta los resultados de los exámenes en la toma de decisiones con relación al ajuste de los programas de las clases
- Hago lo posible para que la puesta en práctica de los programas de estudio sea pertinente
- Cuando un profesor señala un problema que tuvo en clase, buscamos una solución juntos
- Estoy atento a los comportamientos susceptibles de perturbar el trabajo en clase

El rango va de 1 a 4 porque cada ítem puede tomar los valores:

1=Nunca

2=Rara vez

3=Bastante frecuente

4=Muy frecuente

- **Razón de estudiantes por profesor:** Razón de estudiantes por profesor, su rango va de 1.288 a 43.478
- **Tipo de colegio:** La variable tipo de colegio se modificó para formar cuatro variables dummy, tomando como referencia los colegios privados. La distribución es la siguiente:
 - **Tipo de colegio_Público Académico:** Variable dummy que identifica a los colegios privados: 1= Si es un colegio público académico diurno
0= Otro caso

- **Tipo de colegio_Público Técnico:** Variable dummy que identifica a los colegios públicos técnicos diurnos:
1= Si es un colegio público técnico diurno
0= Otro caso
- **Tipo de colegio_Nocturno:** Variable dummy que identifica a los colegios nocturnos:
1= Si es un colegio nocturno
0= Otro caso
- **Tipo de colegio_Rural o Telesecundaria:** Variable dummy que identifica a los colegios rurales o telesecundarias:
1= Si es un colegio rural o una telesecundaria
0= Otro caso
- **Zona del colegio:** Zona del colegio, toma los valores:
 - 1= Urbano
 - 0= Rural
- **IDS2007:** Índice de Desarrollo Social Distrital, su rango va de 0 a 100.

Resultados

El Cuadro 7 da cuenta de los principales hallazgos en la exploración de factores del estudiante y del centro educativo y director que predicen el rendimiento en la prueba PISA de Competencia Lectora.

Debe indicarse que el valor B en el cuadro es el coeficiente de regresión no estandarizado, que está dado en las mismas unidades de medida que la variable dependiente. El coeficiente beta, por el contrario, está libre del efecto de las unidades de medida, de hecho es el coeficiente de correlación parcial entre la respectiva variable independiente y la variable dependiente. Por esta razón, los valores beta pueden compararse directamente entre sí para establecer la importancia relativa de cada variable independiente en la predicción. Además, cada uno de los coeficientes indica el grado de asociación entre la variable dependiente y la respectiva variable independiente, controlando o “manteniendo constantes” las otras variables independientes en el modelo.

Utilizamos como “regla de dedo” el criterio de juzgar importante, a nivel práctico, un valor de beta igual o mayor a 0,10 en valor absoluto. Debe notarse también que la importancia práctica no es lo mismo que significancia estadística. La importancia práctica está asociada al grado en que un resultado se juzga como relevante desde el punto de vista del investigador, considerando el problema y su contexto. La significancia estadística se refiere al grado en que los resultados de la muestra se pueden generalizar hacia toda la población.

Para la interpretación de estos resultados es importante recordar que los coeficientes B (no estandarizados) se interpretan de dos maneras diferentes dependiendo del nivel de

medición de la correspondiente variable predictora o variable independiente. Cuando la variable independiente es métrica, es decir, se mide en unidades que permiten afirmar cuanto mayor es un elemento con respecto a otro, el coeficiente B se interpreta como una pendiente, es decir, nos indica en cuantas unidades aumenta o disminuye Y, la variable dependiente, cuando X, la variable independiente aumenta en una unidad, manteniendo constantes o controlando el efecto de las otras variables independientes en el modelo. El signo positivo o negativo nos indica si la relación es directa o inversa, respectivamente.

Por el contrario, cuando la variable independiente es dummy (dicotómica 0-1) el coeficiente B se interpreta como una diferencia de promedios en la variable Y. B representa entonces la diferencia entre el promedio que toma la variable dependiente en la categoría de interés, aquella que toma el valor de 1 en la variable dummy, y el promedio que toma Y en los sujetos que pertenecen al grupo o categoría de referencia y a los que se asigna el valor de 0 en la variable dummy, teniendo en cuenta, igualmente, que se trata de una diferencia de promedios ajustada, manteniendo constante el efecto de las otras variables independientes en el modelo.

Para ilustrar estas interpretaciones con ejemplos de nuestro análisis, tomemos en el modelo para predecir competencia lectora (Cuadro 7) la línea de resultados correspondiente a la variable independiente Actitud hacia la lectura. Como se trata de una escala cuyo rango va de 1 a 4, el valor de B de 16,48 indica que al aumentar en una unidad la escala de actitud hacia la lectura el puntaje en Competencia Lectora aumenta en 16,48 puntos, en promedio, y manteniendo constante o aislando el efecto de las otras variables independientes en el modelo.

Por el contrario, si en este mismo Cuadro 7 tomamos la línea de resultados que se refiere a la variable dummy “tipo de colegio_Público Académico”, a partir de su definición se indica que el grupo de referencia son los colegios privados. Entonces el valor estimado de B de -13,94 se interpreta como la diferencia en los promedios de Competencia Lectora en PISA cuando se comparan estudiantes de colegios públicos y privados, controlando o aislando el efecto de las otras variables predictoras en el modelo. El valor negativo significa que el promedio del grupo de referencia (estudiantes de colegios privados en este caso) es mayor que el promedio del grupo de interés (estudiantes de colegios públicos académicos diurnos). Si la diferencia fuera positiva entonces, de manera análoga, el promedio del grupo de interés sería mayor al del grupo de referencia.

Cuadro 6
Resultados para el Modelo de Regresión Multinivel con Puntaje PISA 2009 en Competencia Lectora como variable dependiente ^{al}

Nombre de la variable	Coefficiente B	Valor de p	Coefficiente beta	Significancia e importancia práctica
Grado	21,64	0,00	0,266	**
Lecciones fuera de horario escolar	-51,65	0,00	-0,156	**
Actitud hacia la lectura	16,48	0,00	0,135	**
Índice de Desarrollo Social 2007	0,67	0,00	0,119	**
Percepción de eficacia de estrategias para resumir un texto	12,00	0,00	0,105	**
Actividades que realiza cuando va a la biblioteca	-8,07	0,00	-0,103	**
Percepción de eficacia de estrategias para comprender texto	11,20	0,00	0,099	*
Frecuencia de repeticion	-12,42	0,00	-0,091	*
Tipo de colegio_Público Académico Diurno	-13,94	0,06	-0,087	
Percepción sobre el valor del colegio	11,23	0,00	0,076	*
Frecuencia de lectura en línea	7,30	0,00	0,076	*
Tipo de colegio_ Técnico	-13,80	0,16	-0,070	
Zona del colegio	10,58	0,06	0,065	
Dificultades para desarrollar enseñanza	-6,57	0,14	-0,057	
Cantidad de libros en la casa	3,69	0,01	0,054	*
Frecuencia de lectura por iniciativa propia	4,65	0,02	0,052	*
Tenencia de elementos en casa	5,74	0,02	0,050	*
Tipo de colegio_Nocturno	-24,95	0,14	-0,049	
Clima escolar negativo	-6,28	0,22	-0,049	
Número de estudiantes en clase la Español	0,51	0,04	0,046	*
Frecuencia de realización de tareas de lectura para el colegio	-5,08	0,04	-0,043	*
Edad en primer grado	-4,97	0,01	-0,043	*
Comunidad donde está la institución	-3,24	0,20	-0,043	
Uso de técnicas "memorísticas" al estudiar	5,36	0,01	0,042	*
Padre y madre viven con estudiante	6,33	0,03	0,039	*
Número de lecciones de Español	2,33	0,04	0,037	*
Actividades extracurriculares disponibles para estudiantes de noveno	13,05	0,29	0,034	
Percepción acerca del trato que le dan los profesores	-4,71	0,06	-0,033	
Escala de gestión del director	-7,06	0,27	-0,032	
Frecuencia de lectura de textos para el colegio	-3,46	0,19	-0,027	
Ambiente negativo en clase	-2,42	0,26	-0,019	
Total estudiantes	0,00	0,46	0,019	

Expectativas de los padres acerca de la institución	-2,10	0,22	-0,019
Uso de técnicas "analíticas" al estudiar	-2,06	0,38	-0,018
Tipo de colegio_Liceo Rural o Telesecundaria	-8,01	0,62	-0,015
Uso de datos de rendimiento académico en rendición de cuentas	4,34	0,60	0,015
Sexo	2,16	0,35	0,014
Nivel educativo de la madre	0,46	0,38	0,013
Padres extranjeros	-2,76	0,44	-0,012
Nivel educativo del padre	0,38	0,57	0,011
Tenencia de bienes	-3,17	0,67	-0,011
Tiempo de lectura por placer	0,63	0,64	0,009
Uso de métodos para monitorear la práctica de profesores de Español	2,34	0,80	0,008
Instituciones compitiendo por estudiantes	0,85	0,71	0,008
Asistió a preescolar	-1,43	0,32	-0,008
Porcentaje de estudiantes que repitió de 7 a 9	0,03	0,81	0,007
Percepción sobre técnicas pedagógicas para el estudio de textos	0,85	0,76	0,007
Programa o ciclo	0,95	0,83	0,006
Clasificación de ocupación del padre	-0,39	0,52	-0,006
Percepción sobre técnicas pedagógicas en general	-0,60	0,83	-0,005
Porcentaje de estudiantes que repitió de 10 a 12	0,04	0,86	0,004
Hay biblioteca en el colegio	-0,43	0,50	-0,002
Razón de estudiantes por profesor	0,00	0,14	0,000
Razón de computadoras por estudiante	0,00	0,72	0,000
Constante	115,81	0,01	-

a/ (R²=54,6%, n=3203)

** : dos asteriscos indican que la variable independiente presenta las dos características: significancia estadística e importancia práctica: significancia estadística al mostrar un valor de p, probabilidad asociada a la prueba de inferencia estadística, menor a 0,05, indicando que el coeficiente de la respectiva variable independiente es estadísticamente significativo y por tanto se puede generalizar a toda la población; importancia práctica al poseer un valor de beta (coeficiente) estandarizado igual o mayor a 0,10, valor a partir del cual se considera relevante para la predicción a nivel descriptivo.

* : un asterisco indica que la variable independiente presenta una y solo una de las dos condiciones: un valor de p, probabilidad asociada a la prueba de inferencia estadística, menor a 0,05, indicando que el coeficiente de la respectiva variable independiente es estadísticamente significativo, es decir, se puede generalizar a toda la población, o alternativamente, un valor de beta (coeficiente estandarizado) igual o mayor a 0,10, valor a partir del cual se considera relevante para la predicción a nivel descriptivo.

Fuente: Elaboración propia

El valor de la variancia explicada, casi 55%, para este modelo de regresión multinivel donde se intenta predecir el puntaje en Competencia Lectora es considerable, y refleja, desde el punto de vista de estas autoras, el camino recorrido por PISA en la indagación e identificación de variables relevantes para predecir el rendimiento en sus pruebas.

Aproximadamente un 43% de la variancia de los puntajes se explica por factores asociados al estudiante y un 12% por factores asociados a la institución. Estos

hallazgos son consistentes con los resultados de estudios internacionales presentados antes. Sin embargo, el porcentaje debido a factores institucionales debe interpretarse con cautela ya que el tamaño de la muestra de colegios para el modelo multinivel no es muy elevado (n=163). En el caso de la muestra de estudiantes los valores faltantes en muchas de las variables provocan que la muestra efectiva para el análisis multinivel se reduzca a 3203 observaciones.

De acuerdo con estos resultados, el perfil de los estudiantes con mayores y menores rendimientos en la Prueba de Competencia Lectora se presenta en el Cuadro 7. Debe indicarse que este perfil incluye las variables que resultaron con elevada importancia práctica o que fueron estadísticamente significativas en el modelo multinivel, y que además son consistentes con la teoría en su direccionalidad (es decir, la relación que presentan con el puntaje de la prueba va en la dirección que se esperaría a nivel conceptual).

Precisamente hay algunas variables que no se incluyen en el perfil por presentar asociaciones inesperadas (en la dirección contraria a lo esperado). Este es el caso de la variable “Lecciones fuera del horario escolar”, que presenta un coeficiente que en apariencia va en contra de lo establecido por la intuición y la teoría, es decir, manteniendo constantes todas las otras variables en el modelo, entre mayor número de lecciones extra recibe el estudiante su puntaje en PISA será menor. Solo podemos conjeturar sobre las posibles razones para este comportamiento espurio, una de ellas es que precisamente son los estudiantes de menor capacidad intelectual los que tienden a requerir estos apoyos, también llama a la reflexión en torno a la naturaleza y eficacia de estas lecciones adicionales.

Algo similar sucede con la variable “Actividades que realiza cuando va a la biblioteca” pues su coeficiente es negativo e inverso a lo esperado. Se intentó reagrupar sus ítems para que los indicadores reflejaran una dimensión de trabajo escolar y una dimensión de diversión, sin embargo las correlaciones se mantuvieron negativas. De igual manera, solo podemos conjeturar en cuanto a posibles razones para este comportamiento, que en apariencia, parece contradictorio. Para efectos de este informe consideramos que este es un resultado espurio y no refleja una relación de causalidad. Situaciones análogas se dieron en el caso de las “Número de estudiantes en la clase de Español” y “Frecuencia de realización de tareas de lectura”.

Cuadro 8

Costa Rica: Perfiles de alto y bajo rendimiento en la Prueba de PISA 2009 de Competencia Lectora ^{af}

Perfil de alto rendimiento (factores del estudiante)	Perfil de bajo rendimiento (factores del estudiante)
1- Cursa un grado mayor al que debería cursar de acuerdo con su edad	Cursa un grado menor al que debería cursar de acuerdo con su edad
2- Posee un alto valor en la escala de Actitud hacia la lectura	Posee un bajo valor en la escala de Actitud hacia la lectura
4- Posee un alto valor en el índice de percepción de la eficacia de estrategias para resumir un texto	Posee un bajo valor en el índice de percepción de la eficacia de estrategias para resumir un texto
5- Posee un alto valor en el índice de percepción de eficacia de estrategias para comprender un texto	Posee un bajo valor en el índice de percepción de la eficacia de estrategias para comprender un texto

6- Tiene baja frecuencia de repetición en los niveles previos	Tiene alta frecuencia de repetición en los niveles previos
7- Percibe que el valor que le brinda el colegio es alto	Percibe que el valor que le brinda el colegio es bajo
8- Realiza con mucha frecuencia lectura en línea	Realiza con nula o escasa frecuencia lectura en línea
11- Reporta una cantidad elevada de libros en su hogar	Reporta una cantidad baja de libros en su hogar
12- Reporta una frecuencia elevada de lectura por iniciativa propia	Reporta una frecuencia baja de lectura por iniciativa propia
13- Posee un valor alto en el índice Tenencia de elementos en su hogar	Posee un valor bajo en el índice Tenencia de elementos en su hogar
15- Tenía la edad adecuada o menor al ingresar a primer grado	Tenía sobre-edad al ingresar a primer grado
17- Posee un alto puntaje en la escala de uso de técnicas de memorización para estudiar	Posee un bajo puntaje en la escala de uso de técnicas de memorización para estudiar
18- Ambos padres viven con el estudiante	El estudiante vive solo con uno de los padres u otros familiares
19- Reporta un número relativamente alto de lecciones de Español	Reporta un número relativamente bajo de lecciones de Español
Perfil de alto rendimiento (factores de la institución)	Perfil de bajo rendimiento (factores de la institución)
3- Está ubicado en un distrito con alto IDS	Está ubicado en un distrito con bajo IDS

a/ La numeración refleja el orden de importancia de cada variable de acuerdo a la magnitud de su asociación con el puntaje en la prueba

Fuente: Elaboración propia

El Cuadro 8 muestra, por su parte, los resultados del modelo multinivel para predecir el puntaje en la prueba de Alfabetización Matemática.

Cuadro 8
Resultados para el Modelo de Regresión Multinivel con Puntaje en PISA 2009
Alfabetización Matemática como variable dependiente ^{a/}

Nombre de la variable	Coefficiente B	Valor de p	Coefficiente beta	Significancia e importancia práctica
Grado	18,92	0,00	0,256	**
Sexo	-32,43	0,00	-0,233	**
Lecciones fuera de horario escolar	-48,21	0,00	-0,161	**
Actividades que realiza cuando va a la biblioteca	-10,42	0,00	-0,147	**
Percepción de eficacia de estrategias para resumir un texto	12,91	0,00	0,125	**
Percepción de eficacia de estrategias para comprender texto	10,58	0,00	0,103	**
Dificultades para desarrollar enseñanza	-10,36	0,03	-0,098	*
Tipo de colegio_ Público Académico Diurno	-13,44	0,16	-0,093	
Frecuencia de lectura en línea	7,50	0,00	0,086	*
Tipo de colegio_ Técnico	-14,64	0,19	-0,082	
Índice de Desarrollo Social 2007	0,41	0,07	0,081	
Frecuencia de repetición	-9,25	0,01	-0,075	*
Actitud hacia la lectura	8,03	0,00	0,072	*

Número de lecciones de Matemática	3,30	0,00	0,065	*
Padre y madre viven con estudiante	9,34	0,00	0,064	*
Frecuencia de lectura por iniciativa propia	5,03	0,01	0,062	*
Percepción sobre el valor del colegio	7,80	0,01	0,058	*
Actividades extracurriculares disponibles para estudiantes de noveno	17,96	0,21	0,052	
Uso de técnicas "analíticas" al estudiar	4,58	0,02	0,045	*
Cantidad de libros en la casa	2,64	0,06	0,043	
Tipo de colegio_Nocturno	-19,25	0,28	-0,042	
Frecuencia de realización de tareas de lectura para el colegio	-4,34	0,05	-0,041	
Tenencia de elementos en casa	4,01	0,15	0,038	
Zona del colegio	5,21	0,42	0,035	
Frecuencia de lectura de textos para el colegio	-4,03	0,22	-0,035	
Edad en primer grado	-3,48	0,08	-0,033	
Tipo de colegio_Liceo Rural o Telesecundaria	-15,48	0,36	-0,032	
Tenencia de bienes	-6,80	0,30	-0,025	
Nivel educativo de la madre	0,79	0,12	0,025	
Uso de técnicas "memorísticas" al estudiar	-2,87	0,13	-0,025	
Clima escolar negativo	-2,51	0,60	-0,022	
Porcentaje de estudiantes que repitió de 7 a 9	0,08	0,59	0,021	
Escala de gestión del director	-4,08	0,56	-0,021	
Comunidad donde está la institución	-1,35	0,64	-0,020	
Padres extranjeros	-3,99	0,25	-0,019	
Clasificación de ocupación del padre	-1,05	0,11	-0,016	
Programa o ciclo	2,13	0,63	0,015	
Instituciones compitiendo por estudiantes	1,45	0,61	0,015	
Nivel educativo del padre	0,40	0,48	0,013	
Percepción acerca del trato que le dan los profesores	1,30	0,48	0,010	
Total estudiantes	0,00	0,80	0,008	
Uso de datos de rendimiento académico en rendición de cuentas	2,14	0,82	0,008	
Tiempo de lectura por placer	0,44	0,68	0,007	
Expectativas de los padres acerca de la institución	0,52	0,82	0,005	
Hay biblioteca en el colegio	-1,09	0,11	-0,005	
Asistió a preescolar	-0,11	0,93	-0,001	
Porcentaje de estudiantes que repitió de 10 a 12	0,00	0,99	0,000	
Razón de estudiantes por profesor	0,00	0,52	0,000	
Razón de computadoras por estudiante	0,00	0,93	0,000	
Constante	150,42	0,00	-	

a/(R2=55,8%, n=3265)

** : dos asteriscos indican que la variable independiente presenta las dos características: significancia estadística e importancia práctica: significancia estadística al mostrar un valor de p, probabilidad asociada a la prueba de inferencia estadística, menor a 0,05, indicando que el coeficiente de la respectiva variable independiente es estadísticamente significativo y por tanto se puede generalizar a toda la población; importancia práctica al poseer un valor de beta (coeficiente) estandarizado igual o mayor a 0,10, valor a partir del cual se considera relevante para la predicción a nivel descriptivo.

* : un asterisco indica que la variable independiente presenta una y solo una de las dos condiciones: un valor de p, probabilidad asociada a la prueba de inferencia estadística, menor a 0,05, indicando que el coeficiente de la respectiva variable independiente es estadísticamente significativo, es decir, se puede generalizar a toda la población, o alternativamente, un valor de beta (coeficiente estandarizado) igual o mayor a 0,10, valor a partir del cual se considera relevante para la predicción a nivel descriptivo.

Fuente: Elaboración propia

Aproximadamente un 45% de la variancia de los puntajes se explica por factores asociados al estudiante y un 11% por factores asociados a la institución. Sin embargo, el porcentaje debido a factores institucionales debe interpretarse con cautela ya que el tamaño de la muestra de colegios para el modelo multinivel no es muy elevado (n=163). En el caso de la muestra de estudiantes los valores faltantes en muchas de las variables provocan que la muestra efectiva para el multinivel se reduzca a 3265 observaciones.

El Cuadro 9 muestra el perfil de los estudiantes con mayores y menores rendimientos en la Prueba de Alfabetización Matemática, según los resultados del modelo e incluyendo las variables que son consistentes con la teoría y que presentan valores relevantes en el indicador de importancia práctica y/o significancia estadística.

Cuadro 9

Costa Rica: Perfiles de alto y bajo rendimiento en la Prueba de PISA 2009 de Alfabetización Matemática ^{a/}

Perfil de alto rendimiento (factores del estudiante)	Perfil de bajo rendimiento (factores del estudiante)
1- Cursa un grado mayor al que debería cursar de acuerdo con su edad	Cursa un grado menor al que debería cursar de acuerdo con su edad
2- Es hombre	Es mujer
3- Posee un alto valor en el índice de percepción de la eficacia de estrategias para resumir un texto	Posee un bajo valor en el índice de percepción de la eficacia de estrategias para resumir un texto
4- Posee un alto valor en el índice de percepción de eficacia de estrategias para comprender un texto	Posee un bajo valor en el índice de percepción de la eficacia de estrategias para comprender un texto
6- Realiza con mucha frecuencia lectura en línea	Realiza con nula o escasa frecuencia lectura en línea
8- Tiene un valor bajo en el índice de repetición escolar	Tiene un valor alto en el índice de repetición escolar
9- Posee un alto valor en la escala de Actitud hacia la lectura	Posee un bajo valor en la escala de Actitud hacia la lectura
10- Reporta un número relativamente alto de lecciones de Matemática	Reporta un número relativamente bajo de lecciones de Matemática
12- Ambos padres viven con el estudiante	El estudiante vive solo con uno de los padres u otros familiares
13- Reporta una frecuencia elevada de lectura por iniciativa propia	Reporta una frecuencia baja de lectura por iniciativa propia
14- Percibe que el valor que le brinda el colegio es alto	Percibe que el valor que le brinda el colegio es bajo
15- Posee un alto puntaje en la escala de uso de técnicas "analíticas" para estudiar	Posee un bajo puntaje en la escala de uso de técnicas "analíticas" para estudiar
Perfil de alto rendimiento	Perfil de bajo rendimiento

(factores de la institución)	(factores de la institución)
5- Posee un valor bajo en el índice de factores asociados a dificultades para desarrollar la enseñanza	Posee un valor alto en el índice de factores asociados a dificultades para desarrollar la enseñanza
7- Está ubicado en un distrito con alto IDS	Está ubicado en un distrito con bajo IDS

a/ La numeración refleja el orden de importancia de cada variable de acuerdo a la magnitud de su asociación con el puntaje en la prueba

Fuente: Elaboración propia

Recomendaciones para acciones de política educativa

Desde la perspectiva de estas autoras y en concordancia con la evidencia suministrada por el análisis multinivel, se presentan las siguientes recomendaciones, basadas en las variables de los perfiles que creemos pueden ser objeto de intervención directa con medidas de política educativa.

Con el propósito de mejorar los niveles de rendimiento en Competencia Lectora se recomienda:

1. Promover en los estudiantes una actitud positiva hacia la lectura.
2. Enseñar explícitamente en el aula las estrategias que son eficaces para resumir un texto.
3. Enseñar explícitamente en el aula las estrategias que son eficaces para comprender un texto.
4. Promover la lectura “en línea”.
5. Promover la lectura por iniciativa propia.
6. Enseñar en el aula estrategias apropiadas para el uso de técnicas de memorización.
7. Aumentar el número de lecciones de Español en aquellos colegios donde los estudiantes reportan cantidades más bajas al promedio.

Y para mejorar el rendimiento en Alfabetización Matemática se sugiere:

1. Enseñar explícitamente en el aula las estrategias que son eficaces para resumir un texto.
2. Enseñar explícitamente en el aula las estrategias que son eficaces para comprender un texto.
3. Promover la lectura “en línea”.
4. Promover en los estudiantes una actitud positiva hacia la lectura.
5. Aumentar el número de lecciones de Matemática en aquellos colegios donde los estudiantes reportan cantidades más bajas al promedio.
6. Promover la lectura por iniciativa propia.
7. Enseñar en el aula estrategias apropiadas para el uso de técnicas analíticas al estudiar.
8. Promover en el centro educativo condiciones que generen menos dificultades para desarrollar la enseñanza, remediando aspectos como la falta de profesores calificados y la escasez o inadecuación del material pedagógico, entre otros.

Por otra parte, el hecho de que en ambos modelos el IDS del distrito en que está ubicado el colegio resultara relevante como predictor llama a la reflexión en cuanto a la temática de equidad en términos geográficos y de políticas de desarrollo, pues solamente por residir en un distrito con bajo IDS es más probable para un estudiante obtener valores relativamente bajos en las pruebas de PISA. Esta circunstancia indudablemente está asociada al tema de las diferencias en oportunidades para aprender que posiblemente se presentan según zonas geográficas y merece también recibir atención por parte de investigadores y tomadores de decisiones.

La importancia de las dimensiones de lectura para predecir el rendimiento en ambas pruebas da origen a una rica discusión teórica y aplicada en torno a esta temática, y es muy consistente con los resultados de investigaciones realizadas por PISA (OECD, 2010).

La relevancia de la lectura y su contracara, la escritura, sumada al hecho de que ambas destrezas no se desarrollan naturalmente sino que necesitan ser adquiridas por medio de la enseñanza sistemática, son una razón de la existencia de las pruebas PISA. Así lo explicitan sus promotores y diseñadores.

El dominio de la lectura es la base para el éxito en otras materias y contribuye a la participación plena en la vida adulta. La capacidad para transmitir información por escrito y de viva voz es uno de los mayores activos de la humanidad. El descubrimiento de que la información se puede compartir en el tiempo y en el espacio, más allá de las limitaciones de la propia voz, el tamaño de un lugar y la precisión de la memoria ha sido esencial para el progreso humano. No obstante, aprender a leer y a escribir requiere un esfuerzo, porque no se puede conseguir sin dominar una serie de habilidades complejas. Aunque el cerebro está preparado biológicamente para adquirir el lenguaje, la escritura y la lectura son logros relativamente recientes en la historia del hombre. Por tanto, para ser lectores competentes son imprescindibles la práctica y la dedicación. (OCDE-Santillana, 2011, p. 18)

Para llamar aún más la atención en torno al tema de la importancia de emprender acciones de política educativa enfocadas al desarrollo de las dimensiones de lectura se ejecutó un estudio de simulación en donde se utilizó el modelo de regresión multinivel para predecir cuál sería el rendimiento en PISA de los estudiantes de colegios públicos, si manteniendo sus valores promedio en todas las otras variables, exhibieran promedios iguales a los estudiantes de los colegios privados en las variables de lectura listadas arriba. Los resultados se resumen en el siguiente cuadro.

Cuadro 11

Simulación de los desempeños de estudiantes de colegios públicos en PISA si tuvieran los niveles promedio en las variables de dimensiones de lectura que poseen los estudiantes de colegios privados

Constructo	Simulación			Valores reales		
	Públicos	Privados	Diferencia	Públicos	Privados	Diferencia
Competencia Lectora	467,03	492,96	25,93	398,63	478,26	-79,63
Alfabetización Matemática	410,38	435,78	25,40	432,02	519,74	-87,72

Fuente: Elaboración propia con datos de Walker (2011) y de OCDE-Santillana (2011).

A partir de esta simulación es claro que solamente a partir de la modificación de los niveles en las dimensiones de lectura en los colegios públicos las diferencias monumentales existentes entre estudiantes de colegios públicos y privados en los desempeños de PISA disminuirían dramáticamente. La disminución sería más de un 67% de las diferencias observadas actualmente en el caso de Competencia Lectora y más de un 70% de las diferencias actuales en el caso de Alfabetización Matemática.

Los resultados son contundentes y claros. El mayor desafío que estos hallazgos presentan para el país es la necesidad de iniciar, de inmediato, acciones de política educativa dirigidas a promover asertivamente las dimensiones de lectura en los colegios públicos. Es evidente, a partir de estos rigurosos análisis, que los beneficios educativos derivados de estas acciones serán abundantes.

El otro desafío que indirectamente nos presenta PISA es el relacionado con el bajo nivel de representatividad de estos resultados, que cubren únicamente a un 53% de los jóvenes de 15 años. Esta situación obliga al país a establecer un compromiso imperioso para atender las necesidades educativas de ese otro 47% de jóvenes en este rango de edad que no están representados en estos análisis, y que son, posiblemente, los más vulnerables y desventajados.

Bibliografía

- Asús N. (2013). Nutrición y rendimiento escolar. Argentina: Fundación Ahdonay.
- Barber, M. & Mourshed, M. (2008). Cómo hicieron los sistemas educativos con mejor desempeño del mundo para alcanzar sus objetivos. Chile: PREAL, McKinsey & Company.
- Blanco, R.; Aguerrondo, I.; Calvo, G.; Cares, G.; Cariola, L.; Cervini, R.; Dari, N.; Fabara, E.; Miranda, L.; Murillo, J.; Rivero, R.; Román, M. & Zorrilla, M. (2008). Eficacia escolar y factores asociados en América Latina y el Caribe. Chile: UNESCO y LLECE.
- Bond, T. & Fox, C. (2001). Applying the Rasch model: fundamental measurement in the human Sciences. Mahwah, New Jersey: LEA.
- Carballo, M. (2005). Análisis de los resultados obtenidos en estudios de eficacia escolar en México, comparados con los de otros países. México: REICE, Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, Vol.3, No.2.
- Castejón, A. (2011). El informe PISA 2009. OCDE. En: Revista Española de Educación Comparada, N° 18, 2011.
- Catalano, G. (2010), Investigating school autonomy: a comparison between England and Italy. Politecnico di Milano, Department of Management, Economics and Industrial Engineering: AEDE.
- Clavel, J. & Balibrea, J. (2010). Motivación y rendimiento académico: los intangibles de la educación. Universidad de Murcia, Departamento de Métodos Cuantitativos para la Economía, Facultad de Economía y Empresa: AEDE.
- Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. (1966). Equality of Educational Opportunity. Washington, D.C.: Government Printing Office.
- Coneus, K.; Laucht, M. & Reub, K. (2010). The role of parental investments for cognitive and non cognitive skill formation – Evidence for the first 11 years of life. Germany, Centre for European Economic Research and Central Institute of Mental Health: AEDE.
- Dorans, N. J. & Holland, P. W. (2000). Population invariance and the equability of test: Basic theory and the linear case. *Journal of Educational Measurement*, 37, 281-306.
- Earley, P. & Weindling, D. (2004). Understanding School Leadership. UK: BELMAS-British Educational Leadership Management and Administration Society.

- Eberts, R. & Stone, J. (1988). Student achievement in public schools: Do principals make a difference? *Economics of Education Review*. <http://econpapers.repec.org/article/eeeecoedu/v3a73ay3a19883ai3a33ap3a291-299.htm>
- Eskeland, G. & Filmer, D. (2002). *Autonomy, Participation, and Learning in Argentine Schools: Findings, and Their Implications for Decentralization*. Argentina: World Bank.
- Florence, M.; Asbridge, M. & Veugelers, P. (2008). *Diet Quality and Academic Performance*. Canadá: American School Health Association.
- Gaviria, J.L. & Castro M. (2005). *Modelos jerárquicos lineales*. Madrid: La Muralla.
- Goldstein, H. (1987). *Multilevel Models in Educational and Social Research*. Londres: Griffin.
- Gronlund N.E. & Linn R.L. (1990). *Measurement and Evaluation in Teaching*. New York: Macmillan.
- Hanushek Eric A. and Woessmann Ludger (2009), "Schooling, cognitive skills, and the Latin American Growth Puzzle". Cambridge: National Bureau of Economic Research. <http://www.nber.org/papers/w15066>.
- Hox, J. (2010). *Multilevel Analysis: Techniques and Application*. 2nd edition. Taylor & Francis Group.
- Kreft, I. & De Leeuw, J. (1998). *Introducing Multilevel Modeling*. ISM: Introducing Statistical Methods. Londres: Sage Publications.
- Leithwood, K.; Anderson, S.; Mascall, B.; Strauss, T.; Moore, S.; Seashore, K.; Wahlstrom, K.; Michlin, M.; Gordon M. & Thomas, E. (2010), *Learning from Leadership: Investigating the Links to Improved Student Learning*. Final Report of Research to the Wallace Foundation. University of Minnesota. <http://www.wallacefoundation.org/knowledge-center/school-leadership/key-research/Documents/Investigating-the-Links-to-Improved-Student-Learning.pdf>
- Machin, S. & Pekkarinen, T. (2009). *Gender differences in test scores*. Londres: MySciNet.
- Martin, M.; Mullis, I.; Gregory, K.; Hoyle, C. & Shen, C. (2000). *Effective Schools in Science and Mathematics, IEA's Third International Mathematics and Science Study*. Chapter 2: Factors Associated with School Effectiveness in Science and Mathematics. Estados Unidos: International Association for the Evaluation of Educational Achievement (IEA). http://pirls.bc.edu/timss1995i/TIMSSPDF/T95_EffSchool.pdf

- Martínez, M. R., Hernández M.J. & Hernández, M.V. (2006). *Psicometría*. Madrid: Alianza Editorial.
- Mella, O. & Ortiz, I. (1999). Rendimiento escolar. Influencias diferenciales de factores externos e internos. Distrito Federal, México: Revista Latinoamericana de Estudios Educativos, 1° trimestre, año/vol. XXIX, número 001, Centro de Estudios Educativos.
- Ministerio de Educación Pública de Costa Rica (2010). Informe Nacional Pruebas Nacionales Diagnósticas de II Ciclo de la Educación General Básica, 2008. San José, Costa Rica: Dirección de Gestión y Evaluación de la Calidad, Departamento de Evaluación Académica y Certificación, Ministerio de Educación Pública.
- Ministerio de Educación Pública de Costa Rica (2012). Informe Nacional de Factores Asociados al Rendimiento Académico en las Pruebas Nacionales Diagnósticas, III Ciclo de la Educación General Básica, 2010. San José, Costa Rica: Dirección de Gestión y Evaluación de la Calidad, Departamento de Evaluación Académica y Certificación, Ministerio de Educación Pública.
- Montero, E (2008). *Escalas o Índices para la medición de constructos: El dilema del analista de datos*. Avances en Medición, 6, 15–24. Bogotá, Colombia: Laboratorio de Psicometría, Facultad de Ciencias Humanas de la Universidad Nacional de Colombia.
- Montero, E. (2001). *Educación e ingreso como predictores de la esperanza de vida: Evidencias de un análisis de regresión aplicado a indicadores de desarrollo humano*. Revista de Ciencias Sociales de la Universidad de Costa Rica: "Individuo y orden social", #94 (diciembre 2001).
- Montero, E. (2001). *La teoría de respuesta a los ítems: una moderna alternativa para el análisis psicométricos de instrumentos de medición*. Revista de Matemática: teoría y aplicaciones. Centro de Investigaciones en Matemática pura y aplicada (CIMPA) y la Escuela de Matemática de la Universidad de Costa Rica. Vol. 7, # 1-2, págs. 217-228.
- Montero, E.; Villalobos, J. & Valverde, A. (2007). *Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico en la Universidad de Costa Rica: un análisis multinivel*. RELIEVE, v. 13, n. 2. http://www.uv.es/RELIEVE/v13n2/RELIEVEv13n2_5.htm.
- Moreira, T.E. (2009). *Relación entre factores individuales e institucionales con el rendimiento en matemática: Un análisis multivariado*. Avances en Medición, 7, 115-128. Bogotá, Colombia: Laboratorio de Psicometría, Facultad de Ciencias Humanas de la Universidad Nacional de Colombia.
- Mullis, I.; Martin, M.; Ruddock, G.; O'Sullivan, Ch. & Preuschoff, C. (2011). TIMSS 2011 Assessment Frameworks. Boston: International Association for the Evaluation of

Educational Achievement (IEA).
<http://www.education.gov.za/LinkClick.aspx?fileticket=Ub4vJ%2BeV9ds%3D&>.

- Muñiz, J. (2003). *Teoría Clásica de los Tests*. Madrid: Ediciones Pirámide, S.A.
- Murillo, J. (2004). *Un marco comprensivo de mejora de la eficacia escolar*. México: *Revista Mexicana de Investigación Educativa*, abril-junio, Vol.IX, número 021.
- Nunnally, J.C. & Bernstein, I.J. (1995). *Teoría psicométrica* (3ª ed). México, D.F.: Editorial McGrawHill Latinoamericana.
- OCDE (2010a). *PISA 2009 Results: Overcoming Social Background: Equity in Learning Opportunities and Outcome*. Vol. II. <http://www.oecd.org/PISA/PISAproducts/48852584.pdf>
- OCDE (2010b). *PISA 2009 Results: Learning to Learn: Student Engagement, Strategies and Practice*. Vol. III. <http://www.oecd.org/PISA/PISAproducts/48852630.pdf>
- OCDE (2010c). *PISA 2009 Results: Results: What Makes a School Successful? Resources, Policies and Practices*. Vol. IV. <http://www.oecd.org/PISA/PISAproducts/48852721.pdf>
- OCDE (2010d). *PISA 2009 Results: Learning Trends. Changes in Student Performance Since 2000*. Vol. V. <http://www.oecd.org/PISA/PISAproducts/48852742.pdf>
- OCDE (2010f). *The High Cost of Low Educational Performance*. <http://www.oecd.org/PISA/PISAproducts/PISA2006/44417824.pdf>
- OCDE (2010g). *Pathways to Success: How Knowledge and Skills at Age 15 Shape Future Lives in Canada*. <http://www.oecd.org/PISA/PISAproducts/PISA2006/44574748.pdf>
- OCDE-Santillana (2011). *Informe PISA 2009. Lo que los estudiantes saben y pueden hacer. Rendimiento de los estudiantes en lectura, matemáticas y ciencias*. Vol. I. OCDE, Santillana y Ministerio de Educación de España. Edición en inglés, 2010; en español, 2011.
- OECD (2009). *PISA 2009: Assessment framework – Key competencies in reading, mathematics and science*. Paris: OECD.
- OECD (2011). *PISA in Focus*. 2011/5 Junio. OECD Publishing. <http://www.oecd.org/PISA/PISAINfocus/48789439.pdf>
- OCDE (2011b). *Students On Line. Digital Technologies and Performance* (Volume VI). OECD Publishing.
- OECD (2012), *PISA 2009 Technical Report*, PISA, OECD Publishing. <http://dx.doi.org/10.1787/9789264167872-en>

- Penfield, R. & Camilli, G. (2007). Differential Item Functioning and Item Bias. En S. Sinharay y C.R. Rao (Eds.). *Handbook of Statistics*. Vol. 26, Elsevier.
- Prieto, G. & Delgado A.R. (2003). *Análisis de un test mediante el modelo de Rasch*. Psicothema, vol. 15, nº 1, pp. 94-100.
- Raudenbush, SW. & Bryk, A.S. (2002). Hierarchical Linear Models: Applications and Data Analysis Methods. Second Edition. Series: Advanced Quantitative Techniques 1. Sage Publications: Londres.
- Rojas, L. (2004). Factores Asociados a la Repitencia de los y las Estudiantes que Cursan Séptimo Año en Colegios Académicos, Diurnos y Públicos: Un Análisis de Niveles Múltiples. Tesis doctoral. San José, Costa Rica: Universidad Estatal a Distancia.
- Snijders, T.A.B. & Bosker, R.J. (1999). Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. Londres: Sage Publications.
- Velez, E.; Schiefelbein, E. & Valenzuela, J. (1993). Factores que afectan el rendimiento académico en la educación primaria. Revisión de la Literatura de América Latina y el Caribe. Argentina: Revista Latinoamericana de Innovaciones Educativas.
- Walker, Maurice (2011). PISA 2009 Plus Results. Performance of 15-year-olds in reading, mathematics and science for 10 additional participants. Victoria, Australia: Australian Council for Educational Research Ltd, ACER.
- Zamora, E. (2012), Liderazgo Educativo en Costa Rica, working paper. Alajuela, Costa Rica: INCAE Business School.
- Zancajo, A. & Franquesa, M. (2010). Descomposición del efecto inmigrante en el rendimiento académico en Cataluña según la zona origen. Universidad Autónoma de Barcelona: AEDE.