

Capítulo

5

Rendimiento académico en secundaria: ¿qué aprenden los estudiantes en Costa Rica?

Índice

Hallazgos relevantes	253
Valoración general	255
Introducción	257
Estudios previos sobre factores asociados al rendimiento académico en Costa Rica	258
Pruebas estandarizadas y PISA: una breve caracterización	258
Costa Rica en las pruebas PISA	265
Análisis multinivel sobre el rendimiento académico en las pruebas PISA	269
Análisis multinivel de las pruebas del MEP	272
Modelos de ecuaciones estructurales y relaciones causales: un ejercicio exploratorio con las pruebas PISA	273
Principales desafíos en materia de rendimiento académico	283
Anexo	287

Hallazgos relevantes

- En Costa Rica se han estudiado poco los factores asociados al rendimiento académico en primaria y secundaria.
- A pesar de sus diferencias en alcance y método, los resultados de las pruebas PISA y las pruebas diagnósticas del MEP coinciden al ubicar a la mayoría de los estudiantes en niveles bajos o medios de desempeño con respecto a las habilidades esperadas.
- El desempeño promedio de las y los estudiantes costarricenses en las pruebas internacionales está por debajo del que muestran sus pares de países más avanzados, y no se distinguen particularmente de los resultados obtenidos por otras naciones de América Latina.
- Tanto en Matemática como en lectura existe una marcada brecha entre los alumnos de colegios privados y públicos, y dentro de estos últimos, entre las distintas modalidades de formación. La brecha es más profunda en Matemática. No obstante, ni siquiera la educación privada costarricense se acerca a los niveles de países avanzados.
- La muestra de estudiantes nacionales que participan en las pruebas PISA es menos representativa que en otros países. Por esta razón, se simuló los resultados de Costa Rica para saber si su puntuación promedio sería mayor (o menor) en caso de que la muestra representara a un porcentaje más alto de la población de interés. Se identificó una sobreestimación del puntaje de Costa Rica si se compara con la representatividad de la muestra de Chile. Con respecto a la cobertura de México la posición se mantendría invariable.
- Análisis estadísticos multinivel permitieron identificar los elementos que contribuyen, según el caso, a aumentar o disminuir, el rendimiento académico. Entre ellos destacan factores socioeconómicos asociados a la familia del estudiante y las expectativas del núcleo familiar, combinadas con la percepción de la eficacia de las técnicas de estudio (para comprender y resumir textos) y la actitud hacia la lectura. También inciden las características del centro educativo, como su gestión y su infraestructura, y el nivel de desarrollo socioeconómico de los distritos.
- Una simulación para saber cuál sería el rendimiento de los estudiantes de colegios públicos en las pruebas PISA, si estos exhibieran promedios iguales a los de los alumnos de colegios privados en Lectura, evidencia que un mejoramiento de las capacidades en esta materia reduciría considerablemente las brechas existentes entre ambos grupos en las pruebas estandarizadas.

Valoración general

El desempeño del sistema educativo costarricense ha sido objeto de diversas investigaciones, particularmente en temas relacionados con el currículo, los enfoques pedagógicos, los métodos de enseñanza, la promoción, la repitencia y la exclusión. Sin embargo, un tema crucial para entender ese desempeño: los factores determinantes del rendimiento de los estudiantes, ha sido bastante menos estudiado. Aun menos explorada es la perspectiva comparada de los resultados nacionales con respecto a otros sistemas educativos del mundo.

La reciente participación de Costa Rica, por primera vez, en las pruebas estandarizadas internacionales del programa PISA, abrió una oportunidad para investigar en profundidad no solo el rendimiento académico, sino también los aspectos que inciden sobre él. Se trata de medir la habilidad de los estudiantes para usar los conocimientos adquiridos en la solución de situaciones o problemas cotidianos, y de explorar los factores sociales, del entorno y de la trayectoria personal, que están asociados a esta habilidad. Esa oportunidad para el análisis es la que motiva los esfuerzos de este capítulo especial.

Los resultados de la prueba PISA de competencia lectora ubican a Costa Rica en una posición intermedia cuando se consideran todos los países participantes (puesto 44 de 74) y en el segundo lugar de la región latinoamericana, superada únicamente por Chile. El desempeño en lectura de los alumnos evaluados se concentra en el rango “regular” o “aceptable”.

En la prueba de competencia matemática el país ocupó el puesto 55 de las 74 naciones participantes, por debajo, en el continente, de Estados Unidos, Uruguay, Chile, México y Trinidad y Tobago. Casi dos tercios de los estudiantes no muestran capacidades

básicas para utilizar las herramientas matemáticas que han aprendido para resolver problemas de contexto.

La respuesta institucional a estos resultados ha sido distinta según el área de conocimiento. Por ejemplo, en Matemática el MEP ha impulsado una comprensiva reforma curricular y de gestión educativa que, en principio, estaría dirigida a mejorar las puntuaciones en futuras ediciones de la prueba. En el ámbito de la competencia lectora los cambios han sido más lentos: en 2013 se aprobaron nuevos programas de Español para los ciclos primero y segundo de la educación general básica, en los cuales se hace énfasis en el desarrollo de la competencia comunicativa y lingüística del estudiantado.

Este capítulo va más allá de la comparación de promedios nacionales. Se ha buscado responder las siguientes cuestiones: ¿qué aprenden las y los estudiantes de secundaria en Costa Rica? y, sobre todo, ¿por qué aprenden así? El análisis estadístico aplicado a los resultados de las pruebas PISA revela que hay actitudes y hábitos de los jóvenes, y características de su contexto inmediato, que son claves para mejorar el rendimiento académico. Entre los primeros destacan la actitud positiva hacia la lectura, la eficacia de las estrategias para entender y resumir un texto y el uso de técnicas analíticas para estudiar. Entre los segundos sobresalen el nivel socioeconómico del hogar y el desarrollo social del distrito donde se ubica el colegio.

El uso de la misma metodología con las pruebas diagnósticas que el MEP aplica a las y los alumnos de noveno año, arroja resultados consistentes con los obtenidos en el caso de PISA, aunque no son enteramente comparables, por diferencias en los factores considerados por cada una de estas evaluaciones. Pese a ello, en las dos pruebas del MEP analizadas –Español y Matemática– se determinó

que hay aspectos del entorno inmediato, como el nivel socioeconómico del hogar y las expectativas familiares sobre el logro del estudiante, que tienen relación directa con el rendimiento académico. Esta evidencia señala la necesidad de proveer mecanismos que compensen las brechas originadas por estos factores externos al sistema educativo.

Este conjunto de hallazgos es relevante porque da pautas no solo sobre la orientación que deben tener los esfuerzos en los próximos años, sino también para establecer prioridades en la aplicación de los nuevos programas de estudio y planes de formación docente.

Finalmente, el análisis realizado permite plantear una serie de desafíos al sistema educativo. El primero es identificar aciertos y desaciertos en las primeras experiencias de participación en pruebas estandarizadas internacionales, con miras a corregir los elementos que impiden obtener mejores resultados. El segundo es asumir el compromiso de aspirar a mejores puntuaciones y posicionar al país en los puestos más altos del *ranking* y, en especial, reducir las brechas entre la educación pública y la privada. El tercero es impulsar cambios a la luz de los resultados obtenidos en las pruebas, con miras a mejorar los procesos de enseñanza y aprendizaje que tienen lugar en el aula. Un cuarto desafío es implementar acciones dirigidas a incidir y compensar los factores individuales asociados al rendimiento educativo, en particular los relacionados con las condiciones aceptables para que los estudiantes aprendan. Si el contexto es adverso o dificulta el proceso de aprendizaje, las autoridades educativas y las instituciones públicas deberían buscar mecanismos para cambiarlo, de tal manera que lo que hoy constituye una amenaza, se transforme en una oportunidad para el éxito académico.

INTRODUCCIÓN

Determinar qué aprenden los estudiantes en el sistema educativo es un asunto medular por dos razones fundamentales. Por un lado, es importante para las y los alumnos, pues los conocimientos, habilidades y destrezas adquiridos son las herramientas que en el futuro les permitirán desenvolverse de una manera digna, autónoma y productiva. Por otro lado, es un tema trascendental para el país, que invierte ingentes recursos en la educación con el objetivo de formar personas productivas, que ejerzan su ciudadanía y participen activamente en la sociedad.

Desde hace décadas la cuestión del rendimiento académico ha estado presente en los debates sobre la calidad y la pertinencia de la educación costarricense. Pese a la riqueza de esa discusión, la escasez de evidencia empírica a menudo la ha circunscrito al plano de la confrontación ideológica o de los postulados teóricos. Este capítulo especial del *Cuarto Informe Estado de la Educación* aporta una nueva perspectiva para el abordaje del tema, precisamente mediante el análisis sistemático y en profundidad de la evidencia disponible en esta materia. El énfasis recae en los resultados de las pruebas aplicadas a los alumnos de noveno año, nivel que hasta el 2011 era el mínimo establecido por la Constitución

como meta obligatoria para los estudiantes.

Este capítulo busca responder la siguiente pregunta: ¿qué aprenden las y los estudiantes al finalizar la educación general básica? De esta interrogante general se derivan otras también relevantes: i) ¿existen factores asociados de manera sistemática al mayor o menor rendimiento académico de las y los alumnos?, ii) cuando los resultados obtenidos en Costa Rica se comparan con los de otros países, ¿cómo se puede calificar el desempeño observado? y iii) ¿qué pueden hacer las autoridades educativas, los maestros y profesores, los padres de familia y los mismos estudiantes, con estos resultados?, ¿hay medidas prácticas que se pueden adoptar en el corto plazo?

Para dar respuesta a estas preguntas se utilizaron dos fuentes de información. La primera son los resultados de las pruebas diagnósticas de noveno año aplicadas por el MEP en 2010 en Matemáticas y Español, dos asignaturas clave en el currículo del sistema educativo. La segunda y más novedosa fuente de información son los resultados de las pruebas del Programa para la Evaluación Internacional de Alumnos de la OCDE, o pruebas PISA, en las áreas de Matemáticas, Lectura y Ciencias, en las que Costa Rica participó por primera vez en 2009. PISA es un instrumento de comparación internacional de gran valor

que ha sido poco empleado en el país, situación que este trabajo procura remediar.

El objetivo fundamental de este capítulo es presentar y discutir los principales resultados obtenidos por Costa Rica en las pruebas PISA 2009 de competencia lectora y competencia matemática, y las pruebas diagnósticas del MEP en Español y Matemática. Para el análisis se usaron procedimientos multivariados que van mucho más allá de los datos descriptivos y los *rankings* ordinales, y permiten explorar los factores que subyacen a la información recabada. A partir de estos hallazgos y sus implicaciones se sugieren acciones concretas de política educativa.

El capítulo se organiza en ocho apartados, incluyendo esta introducción. En el segundo acápite se sintetizan, a manera de prólogo, los estudios realizados en el país sobre los factores asociados al rendimiento académico. El tercero describe las características generales de las pruebas estandarizadas y las pruebas PISA, y pone de relieve la poca experiencia nacional en este ámbito. El cuarto apartado informa sobre la ubicación de Costa Rica en el *ranking* de las pruebas PISA. En las secciones quinta y sexta se identifican los factores individuales y contextuales asociados al rendimiento académico, mediante la aplicación de modelos jerárquicos a los resultados de las

pruebas del MEP y PISA. En séptimo lugar se presentan los hallazgos de un ejercicio exploratorio sobre posibles relaciones de causalidad del desempeño educativo, que se realizó mediante la aplicación innovadora de sistemas de ecuaciones estructurales. En la octava y última sección se resumen los principales desafíos del sistema educativo costarricense a la luz de los resultados de las pruebas PISA y las pruebas diagnósticas del MEP.

Estudios previos sobre factores asociados al rendimiento académico en Costa Rica

Si bien a nivel mundial el estudio de factores asociados al desempeño estudiantil tiene una amplia trayectoria, en Costa Rica las investigaciones que utilizan metodologías rigurosas e internacionalmente aceptadas son escasas y, en general, el tema ha sido poco abordado.

Un estudio sobre las relaciones entre factores individuales e institucionales y la nota en la prueba de bachillerato de Matemática (Moreira, 2009) reporta evidencias de una asociación significativa de esta medida de rendimiento con el historial académico y el género del estudiante, el estado civil de los padres, la escolaridad de la madre, el total de miembros del hogar y el espacio físico. Los resultados se refieren a una muestra probabilística, estratificada y proporcional, de alumnos de colegios públicos académicos diurnos. Se usó un modelo de regresión múltiple.

A su vez, Rojas (2004) utilizó un modelo multinivel para explorar factores asociados a la repitencia en séptimo año de colegios académicos, diurnos y públicos, con una muestra de 1.442 alumnos y 115 docentes de Inglés, Español, Estudios Sociales, Ciencias y Matemática. Encontró relaciones directas con dos variables: mala relación entre los padres y el estudiante, según la percepción de este último, y mala interacción entre profesor y estudiante, según la valoración del primero. Además identificó asociaciones inversas con los siguientes factores: autoestima, motivación, la costumbre de realizar una mayor cantidad de tareas, el contar con ayuda para efectuar esas tareas, el respeto percibido por los estudiantes de sus iguales, la residencia con ambos padres durante el tiempo lectivo,

indicadores de condición socioeconómica del alumno y su familia, e indicadores de la infraestructura y las facilidades del centro educativo, incluyendo el acceso a Internet.

Por otra parte, desde el año 2009 el Departamento de Evaluación Académica y Certificación, de la Dirección de Gestión y Evaluación de la Calidad del MEP, analiza los 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, en las que el conglomerado es el centro educativo. Se trabaja con tres niveles de agregación: estudiante, docente y director. Un cuestionario de contexto que deben completar los participantes de cada nivel provee las variables que luego se relacionan con el rendimiento en las pruebas. El primer estudio de este tipo tuvo como propósito identificar elementos predictores en las pruebas diagnósticas de sexto grado aplicadas en 2008 (MEP, 2010).

Más recientemente el MEP publicó el estudio de factores asociados al rendimiento en las pruebas diagnósticas de noveno año (MEP, 2012). Una de las ventajas de este tipo de investigaciones es que se conceptualizan y miden variables a nivel de aula y de docente, lo que posibilita la identificación de aspectos relacionados con la mediación pedagógica y la dinámica de la clase. Además permiten medir rasgos idiosincráticos del país que pueden no ser relevantes en el ámbito internacional. Asimismo, estas pruebas se realizan en las cinco asignaturas principales (Matemática, Español, Estudios Sociales, Ciencias e Inglés), a diferencia de las pruebas internacionales, que tienden a concentrarse en Lenguaje, Matemática y Ciencias. Por último, los constructos objeto de medición en las pruebas 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 apropiados como termómetro del logro de los aprendizajes académicos formales establecidos en los programas de estudios.

Finalmente, cabe señalar que la participación de Costa Rica en pruebas o estudios internacionales es muy reciente. La

primera experiencia se realizó en el marco del Segundo Estudio Regional Comparativo y Explicativo (Serce), impulsado por la Unesco, el cual se aplicó entre 2004 y 2008 en centros educativos de primaria de América Latina y el Caribe. En el caso de secundaria, en 2008 el MEP gestionó la incorporación del país en el programa de las pruebas PISA y una delegación nacional participó por primera vez en el ciclo denominado PISA 2009+.



>>PARA MÁS INFORMACIÓN SOBRE COSTA RICA EN LAS PRUEBAS PISA 2009, véase Montero et al., 2012, en www.estadonacion.or.cr

Pruebas estandarizadas y PISA: una breve caracterización

Las pruebas estandarizadas son instrumentos de evaluación contruidos con base en rigurosos procesos de validación. Se aplican a un conjunto amplio de individuos (en este caso estudiantes de secundaria) y se califican bajo estrictos protocolos, con la finalidad de garantizar, por un lado, la comparabilidad de las puntuaciones y, por otro, la validez de las inferencias sobre el desempeño de las personas examinadas.

Dado que en el ámbito educativo se emplean definiciones y mediciones de conceptos complejos —tales como rendimiento académico, habilidad intelectual, conocimientos, actitudes, valores y competencias— es necesario utilizar herramientas de la Psicometría que permitan generar indicadores empíricos del grado de validez y confiabilidad de las estimaciones realizadas con los instrumentos estandarizados (Martínez et al., 2006; Nunnally y Bernstein, 1995). En la Psicometría y la medición educativa existe una larga tradición científica que ha dado origen a teorías y métodos para construir y validar las pruebas estandarizadas. Entre los segundos se incluyen aplicaciones piloto para validar los ítems o reactivos y obtener evidencias de su calidad técnica, por medio de modelos como la teoría clásica de los test (TCT) o, más recientemente, el modelo de Rasch¹ (recuadro 5.1).

La calidad de las pruebas estandarizadas es un aspecto medular en materia de evaluación de conocimientos. A diferencia de los exáme-

nes que los docentes aplican periódicamente para evaluar a sus alumnos, en las pruebas estandarizadas no se consideran contextos específicos de aprendizaje o mediaciones pedagógicas, ni las dinámicas particulares que se realizan en el aula. Además hay que tener presente que las pruebas diseñadas por los profesores poseen un valor relativo como medida de rendimiento, ya que carecen de un criterio estandarizado para todos los centros educativos, todos los cursos y todo el cuerpo docente. En el cuadro 5.1 se resumen las principales diferencias entre ambos tipos de pruebas.

Validación de las pruebas estandarizadas

En el proceso de validación de un instrumento de medición estandarizado, la recolección de evidencia empírica involucra diferentes actores y métodos. Uno de los más comunes es la consulta a expertos. No obstante, por lo general ésta se considera insuficiente, por lo que se recurre a otras herramientas para recopilar evidencia de validez sólida y suficientemente creíble. Esto involucra una aplicación piloto, combinada con un análisis psicométrico básico del instrumento y de los ítems que lo

Recuadro 5.1

El modelo de Rasch como herramienta para la evaluación educativa en gran escala

Durante muchos años los educadores y los evaluadores han cuestionado la utilidad de la Psicometría en las pruebas educativas, dado que los puntajes generados a partir del modelo de medición más antiguo y conocido de la teoría clásica de los test (TCT) no proveen información acerca de qué puede o no hacer el estudiante. Además, en la TCT se supone que el error de medición asociado a cada puntaje observado es igual para todos los examinados.

En respuesta a las limitaciones de la TCT han surgido nuevos modelos de medición. Uno de ellos es el modelo de Rasch, que presenta ventajas fundamentales para la interpretación y análisis de los resultados de pruebas educativas, a saber:

- A partir de la puntuación obtenida, permite determinar los logros del estudiante y sus áreas de oportunidad en términos del constructo objeto de medición.
- Genera una estimación del error para cada puntuación obtenida en el constructo de interés. Esto permite adaptar las pruebas a los fines buscados y a las poblaciones meta.

- Permite saber si las respuestas brindadas en el examen sugieren la posibilidad de copia o contestación al azar.
- Se cumple el principio de que diferencias iguales en la puntuación representan diferencias iguales en el constructo.
- A diferencia de otros modelos, siempre es estimable y permite obtener resultados.

La fórmula matemática de este modelo relaciona la probabilidad de responder el ítem correctamente, la diferencia entre el nivel en el constructo o atributo del individuo y la dificultad del ítem. Si el nivel del examinado en el atributo que se desea medir es mayor que la dificultad del ítem, entonces la probabilidad de que responda correctamente es mayor que la probabilidad contraria. Se establece la relación inversa en el caso de que el nivel del examinado en el atributo que se desea medir sea menor que la dificultad del ítem, es decir, la probabilidad de responder en forma correcta será menor que la probabilidad contraria.

Fuente: Rojas, 2012.

Cuadro 5.1

Diferencias entre pruebas estandarizadas y pruebas construidas en el aula

Aspecto	Pruebas estandarizadas	Pruebas en el aula
Resultados de aprendizajes y contenidos medidos	Los resultados y contenidos objeto de medición deben ser comunes para todos los centros educativos. Las pruebas rara vez enfatizan aspectos locales del currículo.	Se consideran aspectos específicos del aprendizaje generado a partir de un determinado proceso de mediación pedagógica y una dinámica particular en el 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 elaborados por especialistas, sometidos a pruebas piloto y seleccionados con base en su efectividad para lograr 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 pruebas estandarizadas, por el poco tiempo del que dispone el docente y la falta de conocimientos especializados.
Confiabilidad	Alta confiabilidad, comúnmente mayor a 0,8 (tanto en la TCT como en el modelo de Rasch).	Por lo general desconocida; puede ser alta si la prueba es cuidadosamente construida. Rara vez se calculan medidas empíricas.
Administración y calificación	Procedimientos estandarizados (uniformes). Se brindan instrucciones específicas. Existen protocolos de administración y calificación.	Se favorecen los procedimientos uniformes, pero hay laxitud. Muchas veces los protocolos no existen o son poco específicos.
Interpretación de puntajes	Los puntajes se asocian a niveles de desempeño. Existen manuales y otras guías que ayudan a usar las pruebas e interpretar sus resultados.	Las comparaciones de los puntajes y las interpretaciones se limitan a la situación específica del centro educativo o el aula.
Número de examinados	Aplicación masiva, usualmente varios cientos o miles.	Aplicación reducida, grupos pequeños en un contexto restringido.

Fuente: Montero et al., 2012, con base en Gronlund y Linn, 1990.

componen. Entre los métodos y modelos de análisis que se utilizan en este proceso figuran los siguientes:

- Análisis de factores exploratorio y confirmatorio.
- Teoría clásica de los test (TCT).
- Teoría de respuesta a los ítems (TRI).
- Modelo de Rasch.
- DIF (funcionamiento diferencial del ítem).
- Equiparación de puntajes.

El análisis de factores, tanto exploratorio como confirmatorio, se refiere a técnicas multivariadas que permiten conocer las dimensiones subyacentes a los datos, requisito fundamental para la gran mayoría de los modelos de medición (Martínez et al., 2006; Nunnally y Bernstein, 1995). Por su parte, la teoría clásica de los test (TCT), el más antiguo y conocido modelo de medición, genera indicadores numéricos de la calidad técnica del instrumento, incluyendo su resultado más importante: el coeficiente alfa de Cronbach, que determina la precisión de la prueba en cuanto a consistencia interna y apunta hacia el grado de estabilidad de los puntajes (Muñiz, 2003).

En lo que respecta al funcionamiento diferencial del ítem, se busca identificar preguntas que, en vez de mostrar 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 y Camilli, 2007). Por su parte, la equiparación de puntajes alude 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 que tienen ítems distintos pero intentan medir el mismo constructo (Dorans y Holland, 2000).

Con los modelos TRI (teoría de respuesta a los ítems) y Rasch se obtienen, por un lado, estimaciones de los parámetros del ítem que son menos dependientes de la muestra de examinados y, por otro, esti-

maciones 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 de la prueba (a diferencia de la TCT, en la cual se asume que el error es constante; Martínez et al., 2006; Montero, 2000). En el 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 muy atractiva a nivel práctico y de interpretación, pues permite evaluar el desempeño del estudiante 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 TCT (Bond y Fox, 2001; Prieto y Delgado, 2003).

Cabe indicar que las pruebas PISA emplean el modelo de Rasch para generar y validar los niveles de desempeño, que a su vez definen la interpretación cualitativa que se realiza a partir de los puntajes obtenidos. Otras pruebas internacionales, como los programas Timss, Serce y Terce, también utilizan el modelo de Rasch, en atención a esta misma propiedad de medición conjunta.



>>PARA MÁS INFORMACIÓN SOBRE
EL MODELO DE RASCH Y SUS APLICACIONES,
véase Rojas, 2012, en
www.estadonacion.or.cr

Las pruebas PISA: alcances, metodología y contenidos

Las pruebas PISA (Programa para la Evaluación Internacional de Alumnos) son exámenes internacionales estandarizados y comparables, construidos y administrados por un consorcio que originalmente estuvo formado por los países miembros de la OCDE, pero se amplió en forma paulatina hasta abarcar a 65 naciones o economías en el 2009. Su aplicación comenzó en el año 2000. Costa Rica participó por primera vez en 2010, y lo hizo en el grupo denominado PISA 2009+, integrado por

diez países que no pudieron asistir a PISA 2009. A nivel latinoamericano también participaron en PISA 2009 México y Chile (ambos miembros de la OCDE), Brasil, Uruguay, Perú, Panamá, Colombia y el estado de Miranda en Venezuela (Walker, 2011). En 2012 Costa Rica volvió a tomar parte en esta iniciativa.

Las pruebas PISA son exámenes “de papel y lápiz” dirigidas a estudiantes de 15 años de edad. Más precisamente, la edad de los examinados debe ser de entre 15 años y 3 meses y 16 años y 2 meses. Los participantes realizan una prueba de dos horas que tiene tres componentes: competencia lectora, competencia matemática y competencia científica. La evaluación incluye una combinación de preguntas de escogencia única y de respuesta abierta, en cada una de las tres áreas. Para la revisión de los ítems de respuesta abierta se requiere la colaboración de calificadores altamente capacitados y un diseño que permita controlar el efecto de estos últimos. En 2009 se agregó un componente de lectura digital, pero ese año Costa Rica no participó. Cada ítem está circunscrito a una de cuatro situaciones: personal, educativa/laboral, pública y científica (OCDE, 2009).

En lugar de enfatizar en contenidos curriculares específicos, estas pruebas miden la habilidad de las y los estudiantes para usar sus conocimientos en la resolución de ítems que presentan situaciones o problemas cotidianos. Cada edición de PISA se concentra en una de las áreas de contenido, brindando siempre información básica, pero más general, sobre las otras dos. En 2009 el mayor peso recayó en la competencia lectora (Walker, 2011). El énfasis permite obtener datos más precisos sobre los componentes del constructo bajo medición, lo cual se logra introduciendo más ítems en la prueba.

En el programa PISA se emplea el término “competencia” para denotar su enfoque en la aplicación práctica de conocimientos y destrezas, a diferencia de otras pruebas internacionales, como “Timss” (tendencias internacionales en el estudio de las Matemáticas y las Ciencias), que tienen vínculos más fuertes con marcos de referencia curriculares y buscan medir el dominio de conocimientos, destrezas y conceptos

específicos. El contenido de las pruebas PISA se extrae de amplias áreas temáticas, por ejemplo espacio y forma para Matemática, en contraste con contenidos más 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 las destrezas y los conocimientos necesarios para participar plena y exitosamente en la sociedad moderna (Walker, 2011; OCDE, 2009).

Las pruebas PISA tienen los siguientes atributos:

- Se concentran en la capacidad del estudiante para utilizar sus habilidades y conocimientos al enfrentar los retos de la vida real. Es decir, tratan de determinar lo que el o la joven sabe y puede hacer con lo que aprendió en el sistema educativo, no si domina y reproduce contenidos curriculares específicos.
- Emplean un concepto innovador de “competencia”, entendida como la capacidad del alumno para aplicar sus conocimientos y habilidades a determinadas áreas disciplinarias, así como para analizar, razonar y comunicarse con eficacia cuando plantea, interpreta y resuelve problemas en diversas situaciones.
- Se orientan hacia la formulación de políticas educativas, al facilitar la vinculación de los resultados obtenidos por las y los estudiantes con los datos sobre sus características y los factores clave que inciden en su aprendizaje, dentro y fuera del centro escolar. De este modo es posible identificar las características de los alumnos, los centros y los sistemas educativos que muestran niveles altos de rendimiento.
- Dan relevancia al aprendizaje a lo largo de la vida, pues no se limitan a evaluar las competencias de los estudiantes en determinados campos disciplinarios, sino que además buscan conocer sus motivaciones, sus opiniones sobre sí mismos y sus estrategias de aprendizaje.
- Dado que se aplican con regularidad, permiten que los países den seguimiento a los avances de sus sistemas educativos.

- Tienen una amplia cobertura geográfica y una metodología colaborativa para su elaboración y aplicación, ya que los expertos de las naciones y economías participantes interactúan con los especialistas internacionales de PISA y entre sí.

- Aplican mecanismos de control de calidad muy estrictos en la concepción, traducción y recolección de muestras de textos y datos, así como en el diseño de las pruebas, lo cual hace que los resultados tengan validez y confiabilidad (OCDE, 2012a).

Estudios longitudinales realizados en Australia, Canadá y Suiza han demostrado que existe una clara relación entre el rendimiento en competencia lectora evaluado por PISA a los 15 años de edad, y posteriores resultados educativos y éxitos en el mundo laboral. Esto sugiere que los niveles de esa competencia predicen con mayor fiabilidad el bienestar económico y social que la educación medida en años de escolaridad. En otras palabras, lo importante es la calidad del aprendizaje, no el tiempo de escolarización (OCDE, 2012a, 2010a y 2010c).

En el caso costarricense, se debe agregar que la participación en este programa no solo proporciona una medida de desempeño de alta confiabilidad y validez, que permite conocer el perfil de los estudiantes de 15 años inscritos en el sistema de educación formal, en cuanto a conocimientos y habilidades para enfrentar el mundo actual, sino que también brinda un marco de referencia conceptual y experiencia acumulada para realizar estudios de factores asociados al rendimiento, que a su vez pueden fundamentar acciones de política educativa.

Como prueba estandarizada de alta calidad técnica, PISA utiliza modelos de medición y análisis “de punta” para generar evidencias sólidas de la validez y confiabilidad de sus resultados, atributos que posibilitan las interpretaciones sustantivas de los puntajes obtenidos en las pruebas (OCDE, 2012a). Además del uso del modelo de Rasch y los métodos DIF (diferencial del ítem), destaca el empleo de complejos modelos de *equating* (equiparación de puntajes) para garantizar que las

diferencias de dificultad entre los distintos cuadernillos o fórmulas de examen y el diferente muestreo de contenidos no afecten las estimaciones del nivel del constructo o la calificación de los examinados.

Competencia lectora y competencia matemática en PISA 2009: resultados para Costa Rica

De las tres áreas evaluadas en las pruebas PISA –competencia lectora, competencia matemática y competencia científica– para este trabajo se decidió analizar los resultados obtenidos por Costa Rica en las dos primeras. La razón principal para incluir la competencia lectora es que fue en este constructo que recayó el énfasis de PISA en el 2009. En el caso de la competencia matemática se tomó en cuenta que en fecha reciente el MEP reformó los programas de estudio en esa asignatura, para enfocarlos, al igual que PISA, en la resolución de problemas; por tal motivo, se consideró muy pertinente una reflexión sustantiva sobre los resultados en ese ámbito. En este acápite se describe el contenido de la evaluación en esos temas, en el primero de ellos a partir del estudio elaborado por Rodino (2012), y en el segundo con base en la investigación de Salas (2012), ambos especialmente comisionados para este Informe.

El *Informe Estado de la Ciencia, la Tecnología y la Innovación*, que será publicado en 2014 por el Programa Estado de la Nación, incluirá un análisis de los resultados en competencia científica. Por ello el tema no se aborda en el presente capítulo.

El constructo “competencia lectora” en las evaluaciones PISA 2009

Las pruebas PISA 2009 y 2009+ describen la competencia lectora como la capacidad individual para “comprender, utilizar textos escritos, reflexionar sobre ellos, e implicarse con ellos para alcanzar los propios objetivos, desarrollar el propio conocimiento y potencial, y participar en la sociedad” (OCDE, 2012a).

La habilidad que concibe PISA va más allá de la tradicional noción de “saber leer”. Representa un concepto amplio y profundo de la lectura. En primer lugar, porque la prueba no indaga si los estudiantes dominan determinados contenidos curriculares,

sino cuánto de lo que aprendieron en el centro educativo pueden aplicar en tareas relacionadas con la vida cotidiana. En segundo lugar, porque además de las pruebas propiamente dichas PISA aplica a los jóvenes un cuestionario muy amplio, que permite analizar sus actitudes, actividades y estrategias de aprendizaje en cuanto a la lectura. Y en tercer lugar, porque no mide la competencia lectora en términos dicotómicos, como algo que una persona tiene o no tiene, sino en forma continua y con base en una estrategia compleja y multidimensional que toma en cuenta varios factores.

Para resolver esta prueba las y los estudiantes deben leer textos muy distintos y responder preguntas sobre ellos. Los materiales que se les presentan corresponden a tres grandes categorías:

- Medio de los textos: incluye textos impresos y digitales, pues la competencia lectora consiste en extraer el sentido del lenguaje verbal expresado en forma gráfica, independientemente del medio que se utilice.
- Formato de los textos: distingue entre textos continuos, discontinuos, mixtos o múltiples. En PISA 2009 y 2009+ se construyeron subescalas para diferenciar textos continuos y discontinuos. En cambio, las dos últimas categorías pueden aparecer en ítems de la prueba, pero no se usan para componer subescalas de evaluación.
- Tipos de texto: descripción, narración, exposición, argumentación, instrucción y transacción. Estos rasgos diferenciadores de los textos se emplean solo para garantizar que la prueba contenga una buena cobertura del material de lectura, pero no para componer subescalas particulares de evaluación.

En la práctica, la competencia lectora se refiere al enfoque cognitivo que adopta una persona al relacionarse con un texto. Se puede definir como el conjunto de estrategias mentales, propósitos o procesos, que utiliza el lector para introducirse en un texto y extraer sentido de él. Se distinguen tres categorías, que son la base para construir otras tantas subescalas de rendimiento:

- Acceder y recuperar: destrezas para encontrar, seleccionar y extraer información de los textos.
- Integrar e interpretar: implica procesar información, a veces entre distintas partes de un texto, para encontrar su sentido. Estas tareas requieren que el lector comprenda las relaciones entre distintas partes del texto, las cuales pueden ser explícitas o no, en cuyo caso requieren una deducción.
- Reflexionar y evaluar: involucra apoyarse en información, ideas o valores externos al texto, incluso en la propia experiencia personal o conocimiento del mundo.

Como se indicó anteriormente, en las pruebas PISA cada uno de los ítems está circunscrito a una de cuatro situaciones: personal, educativa, laboral y pública. El “factor situación” alude al uso que se espera que el lector haga de un texto considerando la intención del autor: uso personal, uso público, uso educativo o uso laboral. Las situaciones escogidas tienen como propósito incorporar un amplio espectro de textos que varían en cuanto a la audiencia y los propósitos para los cuales fueron creados, dándole realismo a la prueba, pero no se emplean para construir subescalas específicas de evaluación. Hay que destacar que PISA toma en cuenta y establece como variables los principales componentes que, según la teoría, interactúan en todo proceso de lectura: texto, lector y tarea (Snow, 2002 y 2010). Del texto se consideran sus rasgos estructurales o de organización; del lector, el enfoque o estrategia que usa para entrar en relación con el texto y darle sentido, y de la tarea, el uso concreto que se espera que el lector haga de distintos textos, a partir del propósito que tenían los autores al escribirlos.

En PISA 2009 se utilizaron más de 130 preguntas de lectura impresa, pero cada participante conoció solo una parte, pues se entregaron diversos conjuntos de preguntas a distintos estudiantes. Las preguntas se organizaron en grupos o cuadernillos que debían ser resueltos en media hora y cada alumno recibió cuatro de ellos, lo que equivale a una prueba de dos horas.

PISA aplica una metodología estándar para crear las escalas de competencia. Las estimaciones que se hacen de la competencia

reflejan los tipos de tarea que se espera que los estudiantes puedan realizar con éxito. Esto significa que es probable que ellos puedan contestar correctamente las preguntas que están en el nivel de dificultad que corresponde a su posición en la escala o por debajo del mismo, pero es poco probable que puedan responder correctamente las preguntas por encima del nivel de dificultad que corresponde a su posición en la escala. Se trata de un modelo probabilístico.

En PISA 2009 y 2009+, el rango de dificultad de las tareas permitió la descripción de siete niveles de competencia lectora, también llamados niveles de desempeño: el nivel 1b es el más bajo y el nivel 6 el más alto². El desempeño mínimo aceptable para la edad de los participantes es el correspondiente al nivel 2 (en el anexo incluido al final de este capítulo se describen en detalle estas categorías).



>>PARA MÁS INFORMACIÓN SOBRE **COMPETENCIA LECTORA SEGÚN PISA 2009+**, véase Rodino, 2012, en www.estadonacion.or.cr

El constructo “competencia matemática” en las evaluaciones PISA 2009

Los conceptos, estructuras y entes matemáticos se han inventado como una herramienta para entender y explicar los fenómenos del mundo natural, social y mental. Más allá de los contenidos matemáticos, en la evaluación PISA lo importante es qué puede hacer el estudiante con esos conceptos y cómo estos le sirven, eventualmente, para comprender las situaciones que enfrenta en su vida.

Se trabaja con ideas principales que engloban hechos y conceptos, los cuales cobran sentido y se encuentran presentes en gran cantidad de situaciones. Esas ideas son:

- Cantidad: implica cuantificar para organizar el mundo. Engloba la comprensión del tamaño relativo, el reconocimiento de las regularidades numéricas y el uso de los números para representar cantidades y atributos cuantificables de los objetos.
- Espacio y forma: el estudio de la forma (percepción espacial) y las construcciones

exige buscar similitudes y diferencias al analizar los componentes formales, y además reconocer las formas en distintas representaciones y en dos o tres dimensiones.

- Cambio y relaciones: las relaciones matemáticas pueden adoptar la forma de ecuaciones o desigualdades, pero también pueden darse relaciones de una naturaleza más general (equivalencia, divisibilidad o inclusión). Se busca evaluar la capacidad de razonamiento funcional, esto es, la capacidad de pensar sobre y en términos de relaciones.
- Incertidumbre: manejo y tratamiento del dato y del azar. Estos dos fenómenos son el objeto de estudio en la estadística y la probabilidad, respectivamente.

Las competencias matemáticas evaluadas por PISA son las siguientes:

- Pensar y razonar: se pretende determinar si el estudiante es capaz de plantearse preguntas como “¿cuántas... hay?”, “¿cómo encontrar...?”, reconocer el tipo de respuestas que ofrece la Matemática para ellas y distinguir entre diferentes tipos de proposiciones (definiciones, teoremas, conjeturas, hipótesis, ejemplos).
- Argumentar: se busca conocer si el estudiante entiende la naturaleza de una prueba matemática y puede seguir y evaluar cadenas lógicas de argumentación, desarrollar procedimientos intuitivos y construir y expresar argumentos matemáticos.

- Comunicar: involucra la capacidad de expresarse en forma adecuada sobre asuntos con contenidos matemáticos, y además la habilidad de comprender las afirmaciones, orales y escritas, de los demás sobre los temas de la disciplina.
- Modelar: busca que el estudiante traduzca la “realidad” en una estructura o modelo matemático y demuestre su capacidad para trabajar con él y validarlo, así como para reflexionar, analizar y plantear críticas a un modelo dado, o los resultados que este aporta, y comunicar eficazmente los resultados obtenidos.
- Plantear y resolver problemas: comprende plantear, formular y definir distintos tipos de problemas matemáticos y resolverlos usando una variedad de métodos aprendidos en el proceso educativo.
- Representar: incluye la capacidad de codificar y decodificar, traducir, interpretar y distinguir entre diferentes tipos de representaciones de objetos y situaciones matemáticas, y escoger entre diversas formas de representación, de acuerdo con una situación y un propósito particulares.
- Utilizar lenguaje y operaciones simbólicas, formales y técnicas: implica la habilidad de decodificar e interpretar lenguaje formal y simbólico, y entender su relación con el lenguaje natural; también supone emplear variables, resolver ecuaciones y realizar cálculos con expresiones algebraicas.

- Usar ayudas y herramientas: involucra el conocimiento y la habilidad de utilizar diversos instrumentos de apoyo que facilitan la actividad matemática (como las tecnologías de información y comunicación o TIC) y además comprender las limitaciones de esas herramientas.

Estas competencias son clasificadas en tres niveles de complejidad, con el fin de que las pruebas incluyan preguntas con distintos niveles de profundidad cognitiva y, de este modo, se logre una evaluación más precisa (cuadro 5.2). A partir de este enfoque se han definido tres grupos de preguntas:

- Preguntas de reproducción: implican la reproducción del conocimiento adquirido; incluyen aquellas que se emplean frecuentemente en las pruebas estandarizadas y en los libros de texto.
- Preguntas de conexión: abarcan las del grupo de reproducción, pero conducen a situaciones en las que la solución de los problemas ya no es simple rutina. No obstante, incluyen escenarios familiares o casi familiares.
- Preguntas de reflexión: incluyen las del grupo de conexión, pero se agrega un elemento de reflexión por parte del estudiante, con respecto a los procesos necesarios o empleados para resolver un problema determinado.

Para efectos de presentar los resultados obtenidos por los participantes, estos tres niveles de complejidad se subdividen a su vez en seis niveles o categorías, que

Cuadro 5.2

Competencias examinadas en las pruebas PISA, según su evaluación en los respectivos niveles de complejidad

Preguntas de reproducción	Preguntas de conexión	Preguntas de reflexión
Representaciones y definiciones estándar	Construcción de modelos	Formulación y solución de problemas complejos
Cálculos rutinarios	Traducción, interpretación y solución de problemas estándar	Reflexión y comprensión a profundidad
Procedimientos rutinarios	Métodos múltiples bien definidos	Aproximación matemática en escenarios más inusuales que los del grupo de conexión
Solución de problemas rutinarios		Múltiples métodos complejos

Fuente: OCDE e INECSE-España, 2004.

muestran el grado de desempeño logrado por los representantes de cada país. El nivel 1 indica que el alumno posee la capacidad de resolver problemas y ejercicios matemáticos rutinarios, y el nivel 6 indica que es capaz de resolver problemas que requieren razonamiento y habilidades matemáticas avanzadas (más detalles en el anexo, al final del capítulo).

Los ítems de la prueba incluyen preguntas de respuesta abierta, respuesta cerrada, respuesta corta, problemas de elección múltiple compleja y preguntas de elección múltiple. Se da preferencia a preguntas cuyos contextos se consideran auténticos y que podrían encontrarse en situaciones reales.

En el desarrollo y elección de los ítems de PISA 2009 se revisó cuidadosamente el nivel de lectura necesario para comprender una pregunta. También se procuró evitar contextos que pudieran comportar un sesgo cultural, y la formulación de las preguntas en la medida de lo posible se realizó en forma natural y directa.

 >>PARA MÁS INFORMACIÓN SOBRE EL CONSTRUCTO "COMPETENCIA matemática" según pisa véase Salas, 2012, en www.estadonacion.or.cr

Clasificación de los puntajes en PISA

Los puntajes en las escalas de PISA representan niveles a lo largo de un *continuum* de destrezas en cada conjunto de competencias. Como se señaló anteriormente, se han establecido rangos para esos niveles de desempeño, asociados a los puntajes, que describen lo que un estudiante típicamente puede hacer en cada nivel (OCDE, 2009). En PISA 2009 se trabajó con seis categorías tanto en competencia lectora como en competencia matemática, con la salvedad de que en la primera el nivel 1 se dividió 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 el

cual el significado de la puntuación se basa en la posición relativa del examinado con respecto a las demás personas que efectuaron la prueba en su grupo de referencia (Gronlund y Linn, 1990; Montero, 2008). La interpretación de criterios es la más apropiada cuando se trata de pruebas educativas con fines diagnósticos, como es el caso de PISA.

Además de la prueba, los estudiantes completan un cuestionario que recoge información sobre distintas variables que podrían estar asociadas a su rendimiento en el examen. El director o directora también llena un cuestionario que indaga acerca de diversas características de su centro educativo y de su propia gestión. En general, la administración de estos cuestionarios tarda entre veinte y treinta minutos.

La combinación de los datos recogidos mediante los cuestionarios y los resultados de las pruebas permite generar estudios de gran utilidad, que amplían el conocimiento de los factores relacionados con el estudiante y la institución que predicen el rendimiento en los exámenes. Además, la identificación de variables asociadas al puntaje en las pruebas sirve de base para formular recomendaciones específicas de política educativa.

Muestreo de la prueba PISA en Costa Rica

Siguiendo los protocolos generales de PISA, en Costa Rica se tomó una muestra aleatoria estratificada de conglomerados que estuvo constituida por 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, a partir del listado general de todos los jóvenes de 15 años, una muestra de treinta estudiantes, independientemente del nivel en que estuvieran matriculados (recuadro 5.2).

Se trabajó con un total de trece cuadernillos de examen distintos, siete de los cuales fueron de menor dificultad que los administrados en los países de mayor desarrollo. Estos se ofrecieron como opción a las naciones que en la escala de lectura habían logrado un puntaje promedio de 450 o menos, o a los participantes nuevos (como Costa Rica) que, de acuerdo con

sus respuestas a una aplicación piloto que se efectuó en 2008, se esperaba que obtuvieran valores en ese rango. La aplicación de 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 (OCDE, 2012a). Por otra parte, mediante un complejo diseño de *equating* (equiparación), se controla el efecto de las diferencias de dificultad en los exámenes, así como en los puntajes y las estimaciones del nivel de desempeño (OCDE, 2012a).

Para minimizar los efectos de posibles sesgos culturales o de contexto, todos los ítems incluidos en los 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 MEP. Se hicieron las modificaciones necesarias para reemplazar aquellos reactivos en los que se sospechó alguna posibilidad de sesgo por cultura, contexto o vocabulario (E³: Mena, 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 de interés no está dentro del sistema educativo, los resultados de PISA presentan un sesgo de sobreestimación, pues los jóvenes que no asisten al colegio o que, aun estando matriculados, no forman parte, por otras razones, del marco muestral de PISA, arrojarían, probablemente, valores más bajos en los puntajes de estas pruebas, comparados con los de los muchachos que sí están dentro de la muestra. Por tanto, es importante indicar que los resultados obtenidos por Costa Rica son representativos de toda la población de 15 años cubierta por el sistema educativo formal, y que no han recibido adecuaciones curriculares; no reflejan los niveles de competencia de todos los jóvenes de 15 años en las áreas o constructos evaluados (Montero et al., 2012).

Recuadro 5.2

La primera experiencia en la aplicación de las pruebas PISA

La primera experiencia de las pruebas PISA en Costa Rica fue exitosa, aunque tuvo algunos inconvenientes que fueron solucionados en el proceso de aplicación. Por ejemplo, en el país la proporción de estudiantes con necesidades educativas especiales (adecuaciones curriculares) ronda el 12%, mientras que a nivel mundial es de 5%, por lo que PISA solo acepta un 5% de exclusión por este motivo. En consecuencia, para el ciclo 2009+ una vez seleccionada la muestra, se valoró caso por caso si el estudiante tenía la capacidad de hacer la prueba y, de lo contrario, se le excluía, teniendo el cuidado de no superar el 5% establecido. En el ciclo 2012 Costa Rica fue el único país de América que aplicó una prueba más corta a esta población.

Como parte del cumplimiento de los requisitos de PISA, en 2010 se recibió la visita de Maurice Walker, director del ciclo 2009+, quien debía corroborar que el MEP tuviera las condiciones necesarias y suficientes para llevar a cabo las pruebas. Luego de entrevistarse con el Ministro de Educación y reunirse con representantes de la Dirección de Gestión y Evaluación de la Calidad (DGEC) del MEP –encargada de aplicar pruebas nacionales e internacionales–, así como con el equipo PISA-Costa Rica, el funcionario dio su visto bueno, por lo que se procedió a desarrollar el cronograma acordado (cuadro 5.3).

En diciembre de 2011 se dieron a conocer los resultados obtenidos por los estudiantes costarricenses en la prueba del ciclo 2009+. Dada la gran cantidad de información recogida mediante la evaluación de PISA, la DGEC decidió divulgarla en una serie de charlas dirigidas a diversas audiencias, entre ellas universidades públicas, funcionarios de las oficinas centrales y las direcciones regionales del MEP y docentes de Matemática.

Existe una gran expectativa de que los resultados de PISA generen cambios en el sistema educativo. Lo cierto del caso es que PISA muestra debilidades y fortalezas, y con fundamento en ellas, las autoridades tienen la posibilidad de propiciar ajustes sustentados en una evaluación que emplea estándares internacionales. Hasta ahora, la principal reforma impulsada por el MEP a partir de esta experiencia son los nuevos programas de Matemática, cuya aplicación se inició en 2013. Cabe señalar que, para tener certeza de la veracidad de los resultados de PISA en Costa Rica, es necesario dar seguimiento a cada uno de ellos en futuras pruebas, por lo que las decisiones basadas en ellos se deben adoptar paulatinamente.

Fuente: Mena, 2012

Cuadro 5.3

Cronograma de aplicación de PISA en Costa Rica

Año	Actividad	Ciclo PISA
2009	Estudio piloto	2009+
2010	Estudio principal	2009+
2011	Reportes e informes	2009+
	Estudio piloto	2012
2012	Estudio principal	2012
2013	Reportes e informes	2012

Fuente: Mena, 2012.

Costa Rica en las pruebas PISA

El análisis de los resultados de Costa Rica en contraste con los obtenidos por el resto de los países evaluados, permite poner en perspectiva el desempeño del sistema educativo nacional y dirigir la mirada hacia los objetivos que se quiere alcanzar en el mediano y largo plazos.

Ubicación en el ranking global

En la escala global de competencia lectora, los estudiantes costarricenses alcanzaron una calificación promedio de 443 puntos. Esto ubica al país en el lugar 44 entre las 74 naciones y economías que participaron en la

prueba. La posición relativa de Costa Rica puede ser evaluada de distintas maneras, según el parámetro que se utilice. Es muy buena si se compara con los demás países latinoamericanos que se sometieron a la prueba, ya que está casi a la par de la nación que obtuvo el mejor puntaje, Chile (449), y por encima de todas las otras: Uruguay (426), México (425), Colombia (413), Brasil (412), Argentina (398), Panamá (371) y Perú (370). Nótese que dos de estos países pertenecen a la OCDE: Chile, con el que Costa Rica tiene escasas diferencias, y México, al que supera.

Sin embargo, la posición de Costa Rica no es tan favorable si se compara con el

resto de los países del continente y del mundo. En el ámbito continental supera a la única nación caribeña que realizó la prueba, Trinidad y Tobago (416), pero está muy por debajo de Estados Unidos (500). En la comparación mundial se ubica a la altura del 40% de la escala (cuadro 5.4).

Como referencia cabe indicar que los primeros lugares del *ranking* fueron ocupados por la ciudad china de Shanghái (556 puntos) y dos países de la OCDE, Corea (539) y Finlandia (536). A nivel mundial, el puntaje promedio de Costa Rica es estadísticamente equivalente a los de Malta, Serbia y Bulgaria (Walker, 2011).

El desempeño en lectura de los estudiantes costarricenses fue “regular” o “aceptable”, ya que la mayoría de los jóvenes evaluados (el 66% de la muestra) puede leer al nivel de competencia 2 o inferior. El nivel 2 constituye la base o rango mínimo en la competencia lectora, en el cual los estudiantes empiezan a demostrar las destrezas que les permitirán participar efectiva y productivamente en la vida laboral, ciudadana y académica (Walker, 2011).

Las mujeres obtuvieron puntajes superiores a los de los hombres en competencia lectora; sin embargo, esta diferencia está entre las menores de todos los participantes en PISA. Costa Rica alcanzó promedios

Cuadro 5.4

Promedio de desempeño en la escala global de lectura y en cinco subescalas, en países seleccionados^{a/}. PISA 2009

País o economía	Escala global de lectura	Subescalas				
		Acceder y recuperar	Integrar e interpretar	Reflexionar y evaluar	Lectura de textos continuos	Lectura de textos no continuos
Shanghái-China	556	549	558	557	564	539
Corea	539	542	541	542	538	542
Finlandia	536	532	538	536	535	535
Hong Kong-China	533	530	530	540	538	522
Singapur	526	526	525	529	522	539
Australia	515	513	513	523	513	524
Noruega	503	512	502	505	505	498
Suiza	501	505	502	497	498	505
Estados Unidos	500	492	495	512	500	503
Alemania	497	501	501	491	496	497
Irlanda	496	498	494	502	497	496
Promedio OCDE	493	495	493	494	494	493
Italia	486	482	490	482	489	476
España	481	480	481	483	484	473
Chile	449	444	452	452	453	444
Costa Rica	443	446	440	443	447	431
Uruguay	426	424	423	436	429	421
México	425	433	418	432	426	424
Miranda-Venezuela	422	415	421	429	424	415
Trinidad y Tobago	416	413	419	413	418	417
Colombia	413	404	411	422	415	409
Brasil	412	407	406	424	414	408
Argentina	398	394	398	402	400	391
Panamá	371	363	372	377	373	359
Perú	370	364	371	368	374	356

a/ De 74 países que presentaron las pruebas.

Fuente: Elaboración propia con base en Walker, 2011.

similares a los de los países de la OCDE con menor puntaje y se encuentra a media desviación estándar de ellos. En general, los resultados en esta materia apuntan a la urgente necesidad de promover el mejoramiento de la lectura en el país (recuadro 5.3).

En cuanto a la prueba de Matemática, la media del puntaje alcanzado por los estudiantes de los países miembros de la OCDE fue de 496, en contraste con los 409 puntos logrados por la representación nacional. Costa Rica ocupó el puesto 55 de los 74 países participantes. En el contexto americano fue

superada por Estados Unidos, Uruguay, Chile, México y Trinidad y Tobago, pero superó la media de Argentina y Brasil (cuadro 5.5).

Un 43,3% de los estudiantes costarricenses mostró un desempeño básico o superior, lo cual indica que apenas comienzan a adquirir las destrezas que les permitirán utilizar las Matemáticas como herramienta fundamental para su futuro. Por sexo, los resultados evidenciaron diferencias significativas, de hasta 25 puntos, a favor de los hombres.

Analizando con mayor detalle los resultados de Costa Rica, se observa que el 56,7% de los alumnos se ubicó en el

nivel 1 o por debajo de él. Esto significa que solo pueden responder preguntas que están claramente definidas, que involucran contextos familiares y en las que todos los datos relevantes están presentes, es decir, solo pueden usar la información necesaria para realizar procedimientos rutinarios, siguiendo instrucciones directas en situaciones explícitas. En particular, la proporción que se encuentra en el nivel 0, que corresponde al 23,6% de la muestra, alude a estudiantes que no tienen éxito en las tareas matemáticas más básicas planteadas en PISA. Su patrón de respuestas hace prever que no

serían capaces o tendrían dificultades para resolver problemas de contexto utilizando las herramientas matemáticas que han adquirido en su proceso formativo.

En el nivel 2 se encuentra un 27,8% de los estudiantes costarricenses y, aunque sigue siendo baja, esta categoría corresponde a jóvenes que tienen la capacidad de interpretar y reconocer situaciones en contextos que requieren una inferencia

directa, extraen información relevante presente en una sola fuente, utilizan algún tipo de representación, emplean los algoritmos básicos y procedimientos convencionales, realizan razonamientos directos y hacen interpretaciones literales de los resultados.

El nivel 3 es el último en que el país obtuvo un porcentaje significativo, con un 12,2%, mientras que el nivel 4 solo lo alcanzó un 3% y los niveles 5 y 6 son

metas para las próximas evaluaciones. En atención a estos resultados, el MEP ha impulsado una reforma curricular y de gestión educativa que se resume en el recuadro 5.4.

Representatividad de la muestra de PISA en Costa Rica

La muestra de estudiantes costarricenses en las pruebas PISA es representativa del 53% de la población de 15 años (Walker,

Recuadro 5.3

Hacia una política nacional de lectura

En diversas ocasiones las autoridades del MEP y del Consejo Superior de Educación (CSE) han señalado la necesidad de que el sistema educativo redoble esfuerzos para promover el desarrollo de las destrezas de lectoescritura, dada su enorme importancia para la vida laboral, social y cívica de las personas. Entre las diez líneas estratégicas del MEP que forman parte del Plan Nacional de Desarrollo 2010-2014 se encuentran varias referencias específicas a este ámbito:

- Estrategia 1: lograr que los estudiantes aprendan lo que es relevante y lo aprendan bien. Contempla la transformación y fortalecimiento de las bibliotecas escolares y la transformación de la enseñanza del Español a lo largo de los ciclos educativos.
- Estrategia 6: elevar en forma sistemática la calidad del recurso humano del sistema educativo. Entre otras acciones propone una capacitación en Español y Matemática vinculada a las transformaciones que se impulsen.
- Estrategia 10: plantea el centro educativo de calidad como eje de la educación costarricense y contempla revisar los objetivos de esta última, distinguiendo entre lo que es realmente relevante y lo accesorio, así como elevar la calidad, la pertinencia y la cobertura.

Esa posición, sin embargo, no se ha traducido en acciones concretas y regulares de promoción de la lectoescritura, más allá de los contenidos curriculares y las lecturas previstas en los programas de primaria y secundaria. La única iniciativa es el llamado Plan Estratégico para el Fomento de la Lectura, acogido por el CSE en su Acuerdo 02-30-05, del 27 de junio de 2005. A este respecto los lineamientos del CSE fueron:

- Establecer para el personal docente la obligación de leer con sus estudiantes, de manera planificada y creativa, durante las lecciones.
- Definir como lecturas obligatorias en clase las lecturas sugeridas que se incluyeron en los programas de Español de primero y segundo ciclos, para el año 2000.
- Solicitar al MEP un plan de acción para implementar este acuerdo, en el que se considere: actualización del personal docente, acceso a la literatura que se ha definido como obligatoria y a otras fuentes literarias, dedicación de dos lecciones semanales a la lectura dirigida y análisis de la situación de las bibliotecas escolares, entre otras.

Adicionalmente, desde finales de 2010 el MEP ha estado trabajando en un documento de planificación denominado “Política Nacional de Lectura y Escritura”, que define catorce prioridades en esta materia, los mandatos para su atención y las instancias responsables. Sin embargo, a la fecha ninguna de estas iniciativas tiene vigencia, pues no han sido nombradas las autoridades correspondientes.

Lo novedoso para el 2013 es la aprobación de nuevos programas de Español para los ciclos primero y segundo de la educación general básica, que tienen como propósito fundamental el desarrollo de la competencia comunicativa y lingüística del estudiantado. Los énfasis son la expresión y la comprensión oral, el disfrute de la lectura y el desarrollo de la escritura de una manera progresiva, con ideas y propósitos claros que permitan a los niños y niñas desenvolverse adecuadamente en diversas situaciones y contextos de la vida cotidiana (MEP, 2013).

Fuente: Rodino, 2012.

Cuadro 5.5

Promedio de desempeño en la escala global de Matemática, en países seleccionados^{a/}. PISA 2009

País	Escala global en Matemática
Shanghái-China	600
Singapur	562
Hong Kong-China	555
Corea	546
China Taipéi	543
Finlandia	541
Suiza	534
Australia	514
Alemania	513
Noruega	498
Promedio OCDE	496
Estados Unidos	487
Irlanda	487
España	483
Italia	483
Uruguay	427
Chile	421
México	419
Trinidad y Tobago	414
Costa Rica	409
Miranda-Venezuela	397
Argentina	388
Brasil	386
Colombia	381
Perú	365
Panamá	360

a/ De 74 países que presentaron las pruebas. Fuente: Elaboración propia con base en Walker, 2011.

2011). En otros países esa cifra es más alta. Además, para el caso nacional los puntajes de las pruebas representan a los jóvenes de 15 años que están cubiertos por el sistema de educación formal y que no han recibido adecuaciones curriculares. Dadas estas condiciones, cabe hacerse dos preguntas. En primer lugar, interesa saber si el puntaje promedio de Costa Rica sería mayor (o menor) si la muestra representara a un porcentaje más alto de la población de interés. En segundo lugar, vale la pena determinar si los cambios en los resultados producto de la simulación alteran significativamente las posiciones del país en el *ranking* de PISA.

Conocida la ubicación de Costa Rica en ese *ranking*, este apartado profundiza en la comparación de los resultados entre países y sus implicaciones prácticas. Para ello se efectuó una simulación de los puntajes de las pruebas basada en el aumento de la representatividad de la muestra nacional, suponiendo que la calificación de los jóvenes que no están representados en la muestra sería menor, por haber tenido, en general, menos oportunidades de aprendizaje.

En el informe de PISA 2009 se propone una metodología para efectuar una simulación como la planteada (OCDE, 2012a):

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%.

Siguiendo esta lógica, la simulación buscó determinar cuáles serían los pro-

medios de Costa Rica en competencia lectora y competencia matemática si, en lugar de tener una cobertura del 53% de la población de interés, esta fuese igual a la que tienen México (61%) y Chile (85%).

Los resultados del ejercicio se presentan en el cuadro 5.6. La diferencia de 32 puntos porcentuales con respecto a Chile produce una sobreestimación de 19,2 unidades en el valor promedio de Costa Rica, 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 produce una sobreestimación del valor promedio de Costa Rica igual a 4,8 unidades = $(8/5) \times 3$.

Para realizar las pruebas de significancia estadística, en las que 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 competencia matemática. En el caso de la simulación con la cobertura de Chile es muy probable que, si se incluyera a la población excluida para aumentar la cobertura al 85%, la muestra provocaría un aumento en la variabilidad de las puntuaciones; entonces se estima, 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.

Aun 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, si Costa Rica exhibiera los porcentajes de cobertura que muestra Chile, sus resultados promedio en ambas pruebas serían significativamente inferiores a los de esa nación.

Por otra parte, en la comparación con México, como se trata de una diferencia de pocos puntos, los errores estándar de los promedios de Costa Rica simulando la cobertura de ese país 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, y las conclusiones son las mismas: en competencia lectora el promedio costarricense es significativamente superior al mexicano, mientras que en competencia matemática sucede lo contrario, o sea, el promedio de México es significativamente más alto que el de Costa Rica.

En síntesis, en cuanto a la primera interrogante planteada al inicio de esta sección, si Costa Rica tuviera la misma cobertura que Chile, sus desempeños promedio en las pruebas PISA serían significativamente inferiores a los de ese país. Cuando se simula la cobertura de México, no hay diferencias significativas en los puntajes.

En lo que concierne a la pregunta de si un cambio en la representatividad de la muestra alteraría las posiciones del país en el *ranking* de PISA, la respuesta es afirmativa

Cuadro 5.6

Costa Rica, Chile y México: comparación entre los promedios obtenidos en las pruebas PISA 2009

Caso	Competencia lectora		Competencia Matemática	
	Promedio	Error estándar	Promedio	Error estándar
Chile	449	3,1	421	3,1
Costa Rica ^{a/}	424	6,0	390	6,0
Costa Rica observado	443	3,2	409	3,0
México	425	2,0	419	1,8
Costa Rica ^{b/}	438	3,2	404	3,0

a/ Simulando para Costa Rica la cobertura de Chile (85%).

b/ Simulando para Costa Rica la cobertura de México (61%).

Fuente: Montero et al., 2012.

Recuadro 5.4

Reforma de la educación matemática en Costa Rica. Avances y desafíos

El 21 de mayo de 2012 el Consejo Superior de Educación (CSE) aprobó nuevos programas de Matemática para los tres ciclos de la educación general básica y el ciclo diversificado. El reto estratégico es aumentar las demandas cognitivas en la educación y potenciar competencias estratégicas mediante acciones en los contenidos, la metodología y la gestión, con la introducción de procesos matemáticos transversales y un trabajo pedagógico equilibrado con problemas de complejidad creciente.

En este sentido, no se pretende aumentar mecánicamente los contenidos cognoscitivos, ni adoptar un currículo “por competencias”, como han hecho otros países. La propuesta más bien busca integrar, a través de la mediación pedagógica, la potenciación de capacidades transversales de mediano y largo plazos (competencias) así como aquellas que se alcanzan en el mediano y corto plazos en las áreas matemáticas (habilidades generales o específicas).

Los planes de estudio se organizan en cinco áreas: números, geometría, medidas, relaciones y álgebra, estadística y probabilidad. La integración vertical de estos programas favorece las conexiones entre ciclos y promueve una visión estratégica de la enseñanza-aprendizaje de esta asignatura. Se incorporan enfoques novedosos, como los siguientes:

- Fortalecimiento del abordaje de la estadística y la probabilidad en todos los años lectivos.
- Estudio de los objetos geométricos a través del movimiento con manipulación dinámica, la visualización espacial y una relación estrecha con el álgebra.
- Mayor énfasis en el desarrollo del sentido numérico, los cálculos y las aproximaciones.
- Introducción temprana y gradual de las relaciones y el álgebra.
- Potenciación del papel de las medidas, tanto en primaria como en secundaria.
- Uso relevante de las tecnologías digitales, con indicaciones precisas.

- Uso de la historia de la Matemática como recurso didáctico.
- Cultivo de actitudes y creencias positivas sobre la Matemática.
- Malla curricular con una estructura novedosa de apoyo al docente, con múltiples indicaciones y ejemplos de método, gestión y evaluación.

La adopción de estos programas se realizará de manera gradual, mediante un plan de transición, entre 2013 y 2016. Al final de ese período estarán incorporados en todos los niveles de la educación académica formal, y quedarán parcialmente pendientes en colegios técnicos profesionales y en la educación abierta, donde el proceso tomará uno o dos años más.

La gradualidad obedece a la profundidad de los cambios realizados, que requieren: i) coherencia con lo que han aprendido los estudiantes hasta ahora, ii) capacitación docente y iii) una oferta de recursos didácticos apropiados.

La implementación exitosa de esta iniciativa plantea algunos retos a los docentes, el MEP y las universidades. Por ejemplo, la incorporación de la estadística y la probabilidad es un desafío, puesto que, si bien el tema forma parte del currículo escolar desde la década de los noventa, no ha sido abordado con la amplitud que le dan los nuevos planes de estudio ni en todos los niveles. Además, no estaba incluido en los temarios de las pruebas de bachillerato y la preparación universitaria de los docentes en esta área ha sido en general deficitaria.

Por otra parte, los nuevos programas proponen, como un eje curricular, el uso de la tecnología. El reto para el país no será tanto disponer de recursos tecnológicos, sino promover su uso inteligente en función de la acción pedagógica en el aula. Lo más relevante será desarrollar una actitud docente alerta y positiva.

El desafío más importante, sin embargo, no se encuentra en los contenidos. Estos programas implican cambios sustanciales en la gestión educativa, la metodología de aula, el planeamiento,

la evaluación cotidiana y la macroevaluación. El esquema predominante en la enseñanza de la Matemática se basa en el modelo de teoría-ejemplo-práctica, pero con la nueva propuesta la idea es empezar con problemas para despertar el interés estudiantil e ir construyendo el aprendizaje de los elementos matemáticos a partir de la situación seleccionada. Además se propone trabajar integralmente, con un solo problema, varias habilidades, y es por ello que el planeamiento, el desarrollo de la lección y la evaluación deben ser distintos.

Una reforma de esta profundidad, que toca todos los ámbitos de la educación nacional, solo podrá llevarse a cabo en el mediano plazo, y tendrá resultados visibles en no menos de una década. La experiencia internacional revela que en los primeros años no se obtendrán mejoras en la promoción estudiantil y la tentación de retroceder será un tema de discusión. El reto es que las autoridades educativas y políticas concedan el tiempo y el apoyo a largo plazo que requiere esta reforma.

En este momento el MEP y el CSE preparan una reforma global al Reglamento de Evaluación de los Aprendizajes, que se espera esté acorde con las necesidades derivadas de los nuevos programas de Matemática. De igual manera, la Dirección de Gestión y Evaluación de la Calidad Educativa del MEP está elaborando una propuesta para incorporar de manera gradual ítems de desarrollo y problemas en las pruebas de bachillerato. Otros departamentos del Ministerio trabajan para ajustar las diversas opciones educativas nacionales a los nuevos planes de estudio.

No obstante, la clave de éxito de la reforma estará en los docentes: la buena preparación que reciban y el compromiso que asuman con ella. Los países que han tenido éxito en mejorar sus sistemas educativos han otorgado una importancia estratégica a los esfuerzos para garantizar la calidad, el profesionalismo y el estatus social de sus docentes.

Fuente: Ruiz, 2012.

en ambas pruebas con la representatividad del caso chileno (en competencia lectora Costa Rica perdería dos posiciones –se situaría por debajo de Uruguay y México– y en competencia matemática perdería una posición y sería superada por el estado de Miranda, en Venezuela). Con la tasa de cobertura de México las posiciones cos-

tarricenses se mantendrían sin variación (Montero et al., 2012).

Análisis multinivel sobre el rendimiento académico en las pruebas PISA

La literatura especializada hace hincapié en los factores individuales y contextuales

como elementos predictores del rendimiento académico, es decir, considera estas dimensiones como dos conjuntos separados uno del otro y le atribuye poca importancia a la combinación de sus efectos. Esta sobresimplificación ha pasado por alto el hecho de que las conclusiones de los estudios “micro” posiblemente no

se sustenten a nivel “macro”, o viceversa. Ignorar los efectos contextuales en estos análisis podría llevar a asumir, en forma errónea, que los efectos estimados afectarían por igual a todos los individuos, o que las diferencias de contexto no son relevantes. En marcado contraste con la mayoría de investigaciones previas, este trabajo pretende subsanar esa debilidad.

En este capítulo se emplea un novedoso enfoque metodológico denominado análisis multinivel, que –como su nombre lo indica– permite combinar niveles de análisis y explorar simultáneamente los efectos de diversas circunstancias. En este caso se examina el impacto de factores sociodemográficos (género, edad, etc.) y contextuales (características de los docentes y los centros educativos), con el objetivo de explicar el rendimiento académico de los estudiantes y determinar en qué medida este obedece a sus rasgos individuales y a su interacción con el contexto en que se desenvuelven.

En la terminología de este tipo de análisis, los aspectos sociodemográficos se consideran factores de nivel 1 y los elementos contextuales son variables de niveles 2 y 3. Este ejercicio se efectúa en tres acápite separados: el primero se enfoca en el constructo “competencia lectora” de las pruebas PISA, el segundo en el constructo “competencia matemática” de la misma evaluación⁴ y el tercero examina las pruebas diagnósticas aplicadas por el MEP a los estudiantes de noveno año.

En todos los casos, se utilizan cuatro modelos estadísticos (uno simple y tres para cada nivel de agregación) y la variable dependiente es el puntaje en la prueba diagnóstica del MEP o la prueba PISA (variable continua). Al primero se le denomina modelo simple, incondicional o de intercepto, y busca determinar si existen diferencias significativas entre colegios en materia de rendimiento académico. Este tipo de modelo permite estudiar la variación entre individuos y entre centros educativos en ausencia de variables a nivel individual y de grupo. En particular, el único efecto aleatorio considerado en este ejercicio es el asociado con el intercepto para cada centro educativo.

A los otros tres modelos se les conoce como modelos con predictores de nivel 1, 2 y 3 (en inglés *random intercept with covariates models*). En ellos se incorporan

las variables en los niveles indicados, las cuales aportan evidencia para identificar, por ejemplo, los factores que determinan las diferencias en el rendimiento académico entre centros educativos. Se consideran una extensión de los típicos análisis de regresión, pues incluyen un intercepto específico para cada institución educativa.

Actitud hacia la lectura, estrategias de estudio y desarrollo social: claves para la competencia lectora

El modelo multinivel aplicado a los factores individuales y contextuales asociados al rendimiento de Costa Rica en la prueba PISA sobre competencia lectora, explica cerca del 55% de la varianza en el puntaje (cuadro 5.7). Este es un nivel explicativo moderadamente satisfactorio. Por su parte, alrededor de un 43% de la varianza de los puntajes se explica por factores asociados al estudiante y solo un 12% por factores asociados a la institución. Estos hallazgos son consistentes con los resultados de estudios internacionales⁵.

Con el afán de simplificar la comprensión de las estimaciones obtenidas con los modelos multinivel, los resultados se sintetizaron de dos maneras. En primer lugar se construyeron gráficos de probabilidades de puntaje en las pruebas, los cuales se complementaron con un perfil de rendimiento. De los primeros se desprende que, entre las variables de nivel 1 (estudiante), la actitud hacia la lectura y la percepción

de eficacia de las estrategias para resumir un texto sobresalen como factores positiva y estadísticamente significativos asociados al rendimiento. Es decir, los estudiantes con una mejor actitud hacia la lectura obtienen en promedio puntajes más altos en la prueba de competencia lectora. De hecho, hay una diferencia de 50 puntos entre los valores extremos del índice (gráfico 5.1). Esto incluye aspectos como:

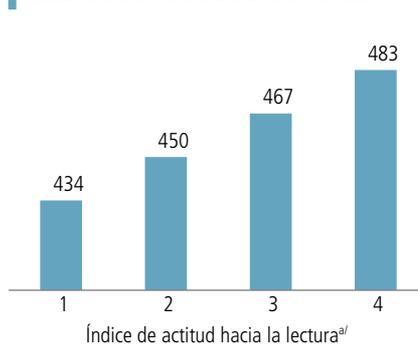
- “La lectura es uno de mis pasatiempos preferidos”
- “Me gusta ir a una librería o una biblioteca”
- “Me gusta intercambiar libros con mis amigos”

Asimismo, los estudiantes que en mayor medida perciben como eficaces las estrategias para resumir un texto (“leo el texto subrayando las frases más importantes”, “elaboro un resumen” y “verifico cuidadosamente si los elementos más importantes del texto aparecen en mi resumen”, entre otras) alcanzan notas superiores que aquellos que las consideran ineficaces (gráfico 5.2).

Finalmente, los puntajes de alumnos que asisten a centros educativos ubicados en distritos con mayores índices de desarrollo social son superiores a los de aquellos que

Gráfico 5.1

Puntaje promedio en la prueba PISA 2009 de competencia lectora, según índice de actitud hacia la lectura^{a/}



a/ Se compone de nueve ítems que miden la actitud del estudiante hacia la lectura, donde 1 es el valor de menor actitud y 4 el de mayor actitud.

Fuente: Elaboración propia con base en Montero et al., 2012.

Gráfico 5.2

Puntaje promedio en la prueba PISA 2009 de competencia lectora, según percepción de eficacia de estrategias para resumir un texto^{a/}



a/ Se obtiene mediante el promedio de la evaluación que da el estudiante a las estrategias para resumir el texto, 1 significa que es una estrategia totalmente ineficaz y 6 significa que es muy eficaz.

Fuente: Elaboración propia con base en Montero et al., 2012.

Cuadro 5.7

Resultado de los modelos multinivel sobre factores asociados a los puntajes en la prueba PISA 2009 de competencia lectora

VARIABLES	Coeficiente	Valor de p	Significancia e importancia práctica ^{a/}
Nivel 1: Estudiante			
Grado	21,64	0,00	**
Edad en primer grado	-4,97	0,01	*
Frecuencia de repetición	-12,42	0,00	*
Padre y madre viven con estudiante	6,33	0,03	*
Tenencia de elementos en casa ^{b/}	5,74	0,02	*
Cantidad de libros en la casa	3,69	0,01	*
Actitud hacia la lectura	16,48	0,00	**
Frecuencia de lectura por iniciativa propia	4,65	0,02	*
Frecuencia de lectura en línea	7,30	0,00	*
Uso de técnicas “memorísticas” al estudiar	5,36	0,01	*
Número de lecciones de Español	2,33	0,04	*
Lecciones fuera de horario escolar	-51,65	0,00	**
Percepción sobre el valor del colegio ^{c/}	11,23	0,00	*
Número de estudiantes en la clase de Español	0,51	0,04	*
Actividades que realiza cuando va a la biblioteca	-8,07	0,00	**
Percepción de eficacia de estrategias para comprender un texto	11,20	0,00	*
Percepción de eficacia de estrategias para resumir un texto	12,00	0,00	**
Frecuencia de realización de tareas de lectura para el colegio	-5,08	0,04	*
Nivel 2: Institución			
Índice de desarrollo social 2007	0,67	0,00	**
Constante	115,81	0,01	

a/ Dos asteriscos (***) indican que la variable independiente presenta dos características: significancia estadística e importancia práctica. La significancia estadística se da cuando la variable muestra un valor de p (probabilidad asociada a la prueba de inferencia estadística), menor a 0,05, lo cual indica que el coeficiente es estadísticamente significativo y, por tanto, se puede generalizar a toda la población. La importancia práctica se determina cuando la variable registra 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 menor a 0,05, o alternativamente, un valor de beta (coeficiente estandarizado) igual o mayor a 0,10.

b/ Índice de tenencia de elementos como televisor, teléfono, computadora, carro y baño con ducha o tina.

c/ Mide el beneficio que, según la percepción del estudiante, le ha brindado el colegio.

Fuente: Elaboración propia con base en Montero et al., 2012.

asisten a colegios localizados en sitios con bajo nivel de desarrollo (gráfico 5.3).

Como ya se indicó, además de los gráficos se elaboró un perfil de los estudiantes con mayores rendimientos en la prueba de competencia lectora. En él se incluyeron las variables que resultaron con elevada importancia práctica o que fueron estadísticamente significativas en el modelo

multinivel, y cuyos efectos se dieron en la dirección esperada. Se excluyeron algunas variables por presentar asociaciones inesperadas (en la dirección contraria a la esperada)⁶. En términos generales, los perfiles de los alumnos con alto rendimiento tienen una relación positiva con factores como hábitos de aprendizaje, prácticas de lectura y exploración, el contexto del

hogar y el desarrollo social del entorno. Los estudiantes de menor rendimiento presentan las características contrarias (cuadro 5.8).

Estrategias para resumir y comprender un texto: factores clave asociados al desempeño en competencia matemática

Los principales hallazgos del análisis de factores determinantes del rendimiento en la prueba PISA de competencia matemática indican que, al igual que en la prueba de competencia lectora, en la explicación predominan los aspectos relacionados con las y los estudiantes. En efecto, en el modelo jerárquico de dos niveles, aproximadamente un 45% de la explicación de la varianza de los puntajes se origina en factores asociados a las personas y tan solo un 11% a elementos propios de las instituciones.

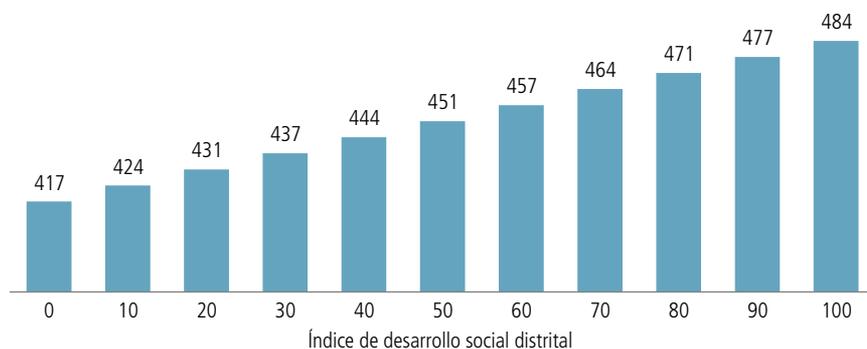
Parte de la explicación de los puntajes en Matemáticas viene dada por aspectos demográficos como la edad (les va mejor a las personas que cursan adelantadas un nivel), el sexo (los hombres obtienen puntajes más altos) y el vivir con ambos padres (cuadro 5.9). Sin embargo, la mayoría de los factores explicativos se relaciona con el historial académico y las prácticas de estudio. Por ejemplo, para un buen resultado en Matemáticas es importante leer, pero más aun que el alumno emplee estrategias adecuadas para comprender y resumir textos. Alcanzan notas superiores quienes consideran eficaz realizar actividades como las siguientes:

- “Leo el texto subrayando las frases más importantes”
- “Elaboro un resumen”
- “Verifico cuidadosamente si los elementos más importantes del texto aparecen en mi resumen”

Esto sugiere interesantes conexiones entre un área (Matemática) y los hábitos generales de aprendizaje de las personas. Factores que afectan negativamente el puntaje en la prueba son, por el contrario, un historial de repetición de un grado y recibir lecciones fuera del horario escolar.

Para explicar la magnitud del impacto de los factores predictores sobre el rendimiento

Gráfico 5.3

Puntaje promedio en la prueba PISA 2009 de competencia lectora, según índice de desarrollo social distrital^{a/}

a/ Este índice es calculado por Mideplan con datos del año 2007. Fluctúa en un rango de 0 a 100, donde 0 corresponde al menor desarrollo social.

Fuente: Elaboración propia con base en Montero et al., 2012.

Cuadro 5.8

PISA 2009: perfiles de alto rendimiento en la prueba de competencia lectora

Tipo de factor	Perfil de alto rendimiento
Factores del estudiante	Cursa un grado mayor al que debería cursar de acuerdo con su edad.
	Valor alto en la escala de actitud hacia la lectura.
	Percibe como eficaces sus estrategias para resumir y comprender un texto.
	Alta frecuencia de repetición.
	Percibe como alto el valor que le brinda el colegio.
	Con mucha frecuencia realiza lecturas en línea.
	Reporta una cantidad elevada de libros en su hogar.
	Reporta una frecuencia elevada de lectura por iniciativa propia.
	Valor alto en el índice de tenencia de elementos en su hogar.
	Tenía la edad adecuada o menor al ingresar a primer grado.
Puntaje alto en la escala de uso de técnicas de memorización para estudiar.	
Ambos padres viven con el estudiante.	
Reporta un número relativamente alto de lecciones de Español.	
Factores institucionales	El colegio está ubicado en un distrito con alto índice de desarrollo social.

Fuente: Montero et al., 2012.

académico, en este apartado nuevamente se recurre a los gráficos de probabilidades y al perfil de desempeño. Se sabe, por lo mostrado en el cuadro 5.9, que la percepción de eficacia de las estrategias para resumir y para comprender un texto, así como el uso de técnicas analíticas para estudiar, sobresalen como variables positiva y significativamente asociadas al rendimiento. En la prueba de competencia matemática,

la diferencia promedio entre quienes consideran eficaz utilizar técnicas para resumir textos y quienes no lo consideran así es de casi 70 puntos (gráfico 5.4).

Asimismo, los estudiantes con mayores puntajes en el índice de eficacia de estrategias para comprender un texto (“subrayo los pasajes más importantes del texto”, “resumo el texto con mis propias palabras” y “leo el texto en voz alta a otra persona”, entre

otras) obtienen en promedio calificaciones más altas en la prueba de competencia matemática. Concretamente, hay una diferencia de 65 puntos entre los valores extremos del índice mencionado (gráfico 5.5).

En el plano de las variables institucionales, el único factor que de modo significativo predice los resultados de la prueba es la existencia de dificultades para desarrollar la enseñanza. Las carencias de personal docente calificado, personal de biblioteca y laboratorios, entre otras, afectan negativamente el puntaje que obtienen las y los estudiantes. De esta manera, los puntajes de quienes asisten a centros educativos con más problemas en este ámbito llegan a ser, en promedio, hasta 30 puntos inferiores que los de quienes se encuentran en la situación contraria. Conforme aumenta el índice de dificultades para la enseñanza disminuye el rendimiento académico (gráfico 5.6).

Con base en los resultados del modelo multinivel, se construyeron los perfiles de los alumnos con mayores rendimientos en la prueba de competencia matemática (cuadro 5.10). Estos son claramente contrastantes con los perfiles de los estudiantes que obtuvieron los menores puntajes. Los factores asociados al bajo rendimiento son aspectos que, en su mayoría, podrían atenderse mediante estrategias específicas impulsadas dentro de los centros educativos.

Análisis multinivel de las pruebas del MEP

Las pruebas diagnósticas que anualmente aplica el MEP a las y los alumnos de noveno año permiten complementar el estudio comparativo internacional de las pruebas PISA. Aunque estas evaluaciones utilizan medidas distintas, es importante determinar si, a pesar de las discrepancias, el análisis multivariado de los factores que inciden sobre el rendimiento estudiantil apunta de modo genérico al mismo tipo de variables. Si bien las diferencias metodológicas impiden llegar a conclusiones sólidas sobre la comparación, ésta permite, al menos, identificar áreas de coincidencia que señalarían derroteros para futuras investigaciones. En particular resulta problemática la adecuada consideración de aspectos relacionados con los hábitos de estudio y lectura en el caso de las pruebas del MEP.

Gráfico 5.4

Puntaje promedio en la prueba PISA 2009 de competencia matemática, según percepción de eficacia de distintas estrategias para resumir un texto



a/ Se obtiene mediante el promedio de la evaluación que da el estudiante a las estrategias para resumir el texto, 1 significa que es una estrategia totalmente ineficaz y 6 significa que es muy eficaz.

Fuente: Elaboración propia con base en Montero et al., 2012.

Gráfico 5.5

Puntaje promedio en la prueba PISA 2009 de competencia matemática, según percepción de eficacia de distintas estrategias para comprender un texto



a/ Se obtiene mediante el promedio de la evaluación que da el estudiante a las estrategias para comprender el texto, 1 significa que es una estrategia totalmente ineficaz y 6 significa que es muy eficaz.

Fuente: Elaboración propia con base en Montero et al., 2012.

Gráfico 5.6

Puntaje promedio en la prueba PISA 2009 de competencia matemática, según índice de dificultades para desarrollar la enseñanza



a/ Valores altos en el índice indican que hay más problemas que afectan la capacidad para desarrollar el proceso de enseñanza.

Fuente: Elaboración propia con base en Montero et al., 2012.

El análisis del rendimiento en las pruebas del MEP también se enfocó en las asignaturas de Matemática y Español y utilizó un modelo multinivel, pero en este caso se trabajó con tres niveles (estudiante, docente/aula y colegio), a diferencia del estudio de las pruebas PISA, en el cual se emplearon dos (estudiante y centro educativo). En 2012 el MEP publicó dos trabajos que sintetizan los principales resultados y hallazgos en torno a esta evaluación (Oviedo, 2012; Muñoz, 2012), los cuales se encuentran disponibles en la página www.estadonacion.or.cr.

Para profundizar el análisis inicial se aplicó la estrategia de definir cuatro modelos. El primero está “vacío” de toda variable, con excepción del centro educativo, a fin

de comprobar si hay diferencias de partida por institución. Luego, sucesivamente, los modelos restantes van incorporando los distintos niveles: el segundo se enfoca solo en las variables de nivel individual, el tercero agrega las del docente/aula y, finalmente, el cuarto incluye las variables de los tres niveles estudiados.

Factores determinantes del rendimiento en la prueba de Español

El primero de los modelos revela diferencias significativas en el rendimiento académico en Español según centro educativo. Esa variabilidad constituye una de las particularidades del desempeño de los estudiantes de secundaria en esta asignatura y justifica el uso de modelos multinivel a

Cuadro 5.9

Resultados de los modelos multinivel sobre factores asociados al puntaje en la prueba PISA 2009 de competencia matemática

Etiqueta de variables	Coficiente	Valor de p	Significancia e importancia práctica ^{a/}
Nivel 1: Estudiante			
Grado	18,92	0,00	**
Sexo	-32,43	0,00	**
Frecuencia de repetición	-9,25	0,01	*
Padre y madre viven con el estudiante	9,34	0,00	*
Actitud hacia la lectura	8,03	0,00	*
Frecuencia de lectura por iniciativa propia	5,03	0,01	*
Frecuencia de lectura en línea	7,50	0,00	*
Uso de técnicas “analíticas” al estudiar	4,58	0,02	*
Número de lecciones de Matemática	3,30	0,00	*
Lecciones fuera de horario escolar	-48,21	0,00	**
Percepción sobre el valor del colegio ^{b/}	7,80	0,01	*
Actividades que realiza cuando va a la biblioteca	-10,42	0,00	**
Percepción de eficacia de las estrategias para comprender un texto	10,58	0,00	**
Percepción de eficacia de las estrategias para resumir un texto	12,91	0,00	**
Nivel 2: Institución			
Dificultades para desarrollar la enseñanza ^{c/}	-10,36	0,03	*
Constante	150,42	0,00	

a/ Dos asteriscos (**) indican que la variable independiente presenta las dos características: significancia estadística e importancia práctica. La significancia estadística se da cuando la variable muestra un valor de p (probabilidad asociada a la prueba de inferencia estadística), menor a 0,05, lo cual indica que el coeficiente es estadísticamente significativo y, por tanto, se puede generalizar a toda la población. La importancia práctica se determina cuando la variable registra 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 menor a 0,05, o, alternativamente, un valor de beta (coeficiente estandarizado) igual o mayor a 0,10.

b/ Mide el beneficio que, según la percepción del estudiante, le ha brindado el colegio.

c/ Valores altos en el índice indican que hay más problemas que afectan la capacidad para desarrollar el proceso de enseñanza.

Fuente: Elaboración propia con base en Montero et al., 2012.

Cuadro 5.10

PISA 2009: perfiles de alto rendimiento en competencia matemática

Tipo de factor	Perfil de alto rendimiento
Factores del estudiante	Cursa un grado mayor al que debería cursar de acuerdo con su edad.
	Es hombre.
	Percibe como eficaces sus estrategias para resumir y comprender un texto.
	Con mucha frecuencia realiza lecturas en línea.
	Valor bajo en el índice de repetición escolar.
	Valor alto en la escala de actitud hacia la lectura.
	Reporta un número relativamente alto de lecciones de Matemática.
	Ambos padres viven con el estudiante.
	Reporta una frecuencia elevada de lectura por iniciativa propia.
	Percibe como alto el valor que le brinda el colegio.
	Puntaje alto en la escala de uso de técnicas "analíticas" para estudiar.
	Valor bajo en el índice de factores asociados a dificultades para desarrollar la enseñanza.
Factores institucionales	Está ubicado en un distrito con alto índice de desarrollo social.

Fuente: Montero et al., 2012.

fin de hallar explicaciones más robustas de ese comportamiento.

En los modelos restantes (2, 3 y 4) sobresalen como buenos predictores del rendimiento en el nivel 1 la edad, el género, la percepción del estudiante sobre la dificultad de la materia, el Español como asignatura de mayor agrado, el índice socioeconómico, la motivación y las expectativas familiares. Más concretamente, los estudiantes de edad promedio (15 años), hombres, que manifiestan que Español es la materia que más les agrada, con mayor nivel socioeconómico, que se sienten motivados por sus familias y perciben altas expectativas sobre su desempeño, muestran mayores puntajes en la prueba de acuerdo con los análisis de significancia estadística. Por el contrario, entre los alumnos de mayor edad y los que consideran que Español es la materia más difícil se observan efectos negativos y estadísticamente significativos entre estos predictores y la variable dependiente (cuadro 5.11).

Por otra parte, se identificaron pocos factores contextuales asociados al rendimiento⁷. Por ejemplo, en el nivel 2, los alumnos de docentes mujeres obtienen puntajes más altos que aquellos que tienen hombres como profesores. Además, a mayor actualización del docente, mayor rendimiento en la prueba. Finalmente, en el nivel 3 los estudiantes de colegios

privados superan a los de centros públicos; de hecho hay una diferencia de 0,6 puntos en la escala original de la prueba entre los dos tipos de instituciones.

Para facilitar la comprensión de las estimaciones mencionadas, se elaboraron gráficos de probabilidades de la puntuación, variando algunas características sociodemográficas como el género, el nivel socioeconómico y las expectativas familiares. En el gráfico 5.7 se observa que, conforme aumenta el índice socioeconómico, mayor es el puntaje proyectado en la prueba de Español. Además, con toda claridad se aprecia que los hombres obtienen puntajes mayores que las mujeres, que esas brechas se mantienen en todos los niveles socioeconómicos y que son estadísticamente significativas a lo largo de todos los niveles.

Asimismo, la evidencia muestra una estrecha relación entre el rendimiento académico y las expectativas familiares sobre el logro educativo. Esto se da particularmente entre las mujeres: tal como ilustra el gráfico 5.8, las estudiantes cuyas familias tienen bajas expectativas en torno a su rendimiento alcanzan puntuaciones menores que las alumnas que se encuentran en la situación contraria. Lo mismo sucede entre los hombres, pero sus calificaciones son mayores que las de las mujeres en las mismas condiciones. Tanto este resultado como el anterior, sobre el nivel socioeconómico, apuntan a la importancia

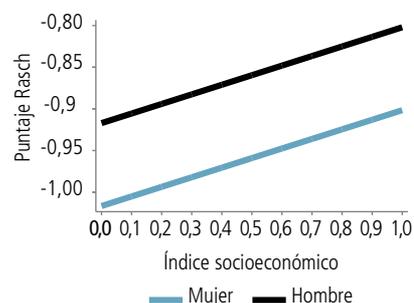
del entorno en el puntaje que obtienen las y los estudiantes en la prueba diagnóstica de Español.

Factores determinantes del rendimiento académico en la prueba de Matemática

Al igual que en la prueba de Español, el primero de los modelos multinivel revela diferencias significativas en el rendimiento académico en Matemática según centro educativo. Existe suficiente variabilidad en los puntajes de las pruebas, condición indispensable para la aplicación de modelos multinivel y la búsqueda de explicaciones a ese comportamiento.

Gráfico 5.7

Puntaje en la prueba diagnóstica de noveno año en Español, según nivel socioeconómico y sexo. 2009



Fuente: Elaboración propia con datos de la Dirección de Gestión y Evaluación de la Calidad del MEP.

Cuadro 5.11

Resultados de los modelos multinivel para la prueba diagnóstica de Español de noveno año. 2009

Variables	Coeficientes ^{a/}			
	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Nivel 1: Estudiante				
Edad		-0,007**	-0,006*	-0,006*
Hombre		0,064***	0,067***	0,066***
Asignatura más difícil		-0,108***	-0,108***	-0,107***
Asignatura que más le agrada: Español		0,182***	0,179***	0,178***
Índice socioeconómico ^{b/}		0,158***	0,154***	0,115**
Motivación familiar		0,079***	0,077***	0,076***
Expectativa familiar del logro del estudiante		0,028*	0,027*	0,026*
Nivel 2: Docente/aula				
Sexo del docente			-0,090**	-0,088**
Actualización de los docentes			0,081**	0,074**
Escala de actividades ^{c/}			0,010	0,006
Nivel 3: Colegio				
Público				-0,560***
Constante	0,074*	-1,059***	-1,255***	-0,706***
Varianza				
Constante	-0,349***	-0,719***	-0,719***	-0,718***
Chi-cuadrado de Wald		432,23	433,24	508,98
Valor de p		0,00	0,00	0,00
Número de casos	5.785	5.551	5.431	5.431

a/ + p<0,10, * p<0,0, **p<0,01, *** p<0,001.

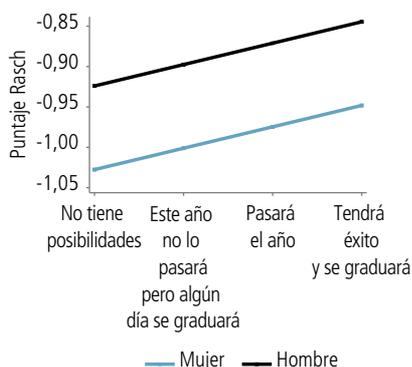
b/ Considera variables como hacinamiento, condiciones de la vivienda y acceso a bienes y servicios.

c/ Actividades realizadas dentro del aula al impartir la lección. La escala refleja la frecuencia con que son utilizadas por el docente.

Fuente: Elaboración propia con datos de la Dirección de Gestión y Evaluación de la Calidad del MEP.

Gráfico 5.8

Puntaje en la prueba diagnóstica de noveno año en Español, según expectativa familiar y sexo del estudiante. 2009



Fuente: Elaboración propia con datos de la Dirección de Gestión y Evaluación de la Calidad del MEP.

En los modelos restantes (2, 3 y 4) se identificaron como factores estadística y positivamente asociados en el nivel 1 la edad, el género, la percepción del estudiante sobre la dificultad de la materia, la Matemática como asignatura de mayor agrado, el índice socioeconómico, la motivación, las expectativas familiares y la percepción sobre el dominio de la materia por parte del docente. En particular, los estudiantes de edad promedio (15 años), mujeres, que manifiestan predilección por la Matemática, con mayor nivel socioeconómico, que se sienten motivados por sus familias, perciben altas expectativas sobre su desempeño y consideran que sus profesores tienen un buen dominio de la materia, obtienen puntajes más altos en la prueba de acuerdo con los análisis de significancia estadística.

En esta prueba se identificó una mayor cantidad de factores de nivel 2 y 3 asociados al rendimiento académico. En el nivel 2, los estudiantes con docentes de mayor edad alcanzan puntajes más altos que los alumnos de profesores menos experimentados. Asimismo, contrario a lo esperado, los alumnos de educadores nombrados en propiedad obtienen notas más bajas que los atendidos por docentes interinos.

En el nivel 3, nuevamente los estudiantes de colegios privados alcanzan puntajes más altos que los de centros públicos. De igual forma, los alumnos de centros diurnos tienen un mejor rendimiento que los de centros nocturnos. Por último, el grado de satisfacción del director, el clima en el colegio y su carácter urbano inciden, aunque de modo marginal, en el rendimiento académico (cuadro 5.12).

Cuadro 5.12

Resultados de los modelos multinivel para la prueba diagnóstica de Matemática de noveno año. 2009

Variables	Coeficientes ^{a/}			
	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Nivel 1: Estudiante				
Edad del estudiante		-0,009*	-0,009+	-0,002
Hombre		-0,053**	-0,045*	-0,043*
Asignatura que más le agrada: Matemática		0,044*	0,055*	0,060**
Índice socioeconómico ^{b/}		0,211***	0,210***	0,137*
Expectativa familiar del logro		0,129***	0,130***	0,126***
Motivación familiar		0,151***	0,163***	0,158***
Índice de actividades ^{c/}		0,119*	0,127*	0,139**
Índice de recursos ^{d/}		0,393***	0,414***	0,401***
Dominio docente		0,068***	0,065***	0,064***
Razones asistencia al colegio ^{e/}		0,169**	0,151**	0,166**
Nivel 2: docente				
Edad del docente			0,008***	0,005+
En propiedad			-0,169***	-0,142**
Tamaño del grupo			0,003	0,006*
Nivel 3: colegio				
Grado de satisfacción del director con la institución				0,121*
Años de experiencia				0,001
Percepción del ambiente escolar				-0,071*
Percepción de los problemas en la institución				0,166*
Urbano				0,101*
Público				-0,553***
Diurno				0,182**
Constante	0,074*	-1,111***	-1,408***	-1,704***
Varianza				
Constante	-0,349***	-0,377***	-0,379***	-0,380***
Chi-cuadrado de Wald	355,67	375,04	574,47	
Valor de p	0,00	0,00	0,00	
Número de casos	5.464	5.150	5.111	

a/ + p<0,10 * p<0,05 ** p<0,01 *** p<0,001

b/ Considera variables como hacinamiento, condiciones de la vivienda y acceso a bienes y servicios.

c/ Actividades que realiza el estudiante para aprender. La escala refleja la frecuencia con que se utilizan.

d/ Herramientas (libros, calculadora, computadora, etc.) usadas por el estudiante. La escala refleja la frecuencia con que se usan.

e/ La variable toma un valor de 1 cuando las razones son académicas y de 0 en el caso contrario.

Fuente: Elaboración propia con datos de la Dirección de Gestión y Evaluación de la Calidad del MEP.

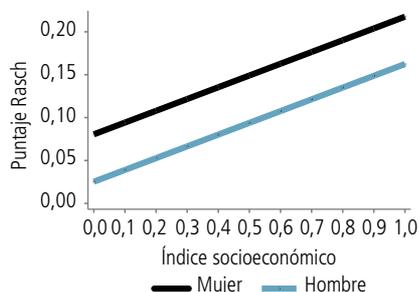
Para simplificar la presentación de los resultados, también en este caso se crearon gráficos de probabilidades de la puntuación, variando las mismas características sociodemográficas: el género, el nivel socioeconómico y las expectativas familiares. En el gráfico 5.9 sobresalen cuatro hallazgos relevantes. En primer lugar, un aumento del índice

socioeconómico del estudiante está asociado a un mayor puntaje en la prueba de Matemática. Además, con claridad se aprecia que las mujeres obtienen mayores puntajes que los hombres, que esas brechas se mantienen en todos los niveles socioeconómicos y que son estadísticamente significativas a lo largo de la mayoría de los niveles.

En cuanto a la relación entre expectativas familiares y rendimiento, la evidencia confirma la asociación positiva entre ambos aspectos. Se identificaron diferencias por género, pero considerablemente menores que las detectadas en la prueba de Español (gráfico 5.10). Así, las estudiantes cuyas familias tienen bajas expectativas en torno

Gráfico 5.9

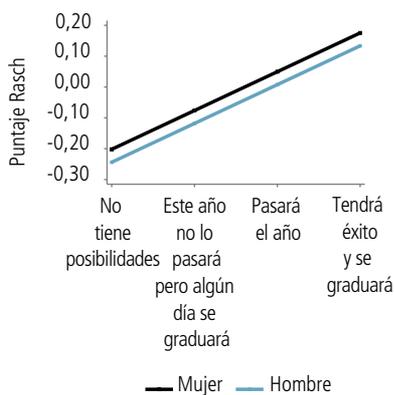
Puntaje en la prueba diagnóstica de noveno año de Matemática, según nivel socioeconómico y sexo. 2009



Fuente: Elaboración propia con datos de la Dirección de Gestión y Evaluación de la Calidad del MEP.

Gráfico 5.10

Puntaje en la prueba diagnóstica de noveno año de Matemática, según expectativa familiar y sexo. 2009



Fuente: Elaboración propia con datos de la Dirección de Gestión y Evaluación de la Calidad del MEP.

a su logro académico obtienen puntuaciones menores en la prueba de Matemática, en comparación con las alumnas que se encuentran en la situación contraria.

Modelos de ecuaciones estructurales y relaciones causales: un ejercicio exploratorio con las pruebas PISA

Los modelos de ecuaciones estructurales o SEM (*structural equations models*) se han hecho populares en los últimos veinticinco años, en virtud de que representan una alternativa para explorar posibles planteamientos teóricos causales, con datos que provienen de estudios *observacionales*

o *correlacionales* (Bollen, 1989; Freedman, 2010)⁸. En la investigación social y educativa resultan muy útiles, pues con frecuencia es imposible, por razones éticas o prácticas, realizar experimentos, que son los diseños óptimos para generar evidencias de causalidad, gracias a que la asignación de los sujetos a los grupos se hace al azar.

En los estudios de observación, por el contrario, se estudia el fenómeno en condiciones “naturales” y, por tanto, es imposible asegurar el control total de “terceras” variables o variables de confusión, que amenazan la validez del argumento causal entre cualesquiera dos variables de interés. Es en este contexto que los SEM ofrecen una propuesta estadística poderosa para representar planteamientos teóricos complejos, con múltiples constructos y con diversas hipótesis acerca de las dependencias (o causalidades) entre ellas.

Los SEM se han utilizado en prácticamente todas las disciplinas científicas, aunque su uso está más difundido en las Ciencias Sociales y del Comportamiento (Bollen, 1989; Freedman, 2010; Kaplan, 2009; Mulaik, 2009). En estudios de factores asociados al rendimiento en pruebas estandarizadas, como PISA, los SEM son una vía para generar evidencias empíricas acerca de posibles planteamientos de causalidad entre las diversas variables y constructos de interés.

En esta sección final del capítulo se presenta un ejercicio preliminar, de carácter ilustrativo, para ejemplificar el potencial de este enfoque analítico en la investigación educativa. Se escogió la prueba PISA de competencia lectora y algunas de las variables del estudiante y del centro educativo que se consideran relevantes para la predicción de sus resultados. Con ellas se desarrollaron diversos ensayos para la estimación de modelos estructurales que tuvieran sentido teórico, que pudieran estimarse de manera correcta desde el punto de vista técnico y que además arrojaran niveles relativamente aceptables de ajuste estadístico. Se utilizó el paquete LISREL, que es el más antiguo y uno de los mejores *software* especializados en SEM.

Por problemas de estimación matemático-estadísticos fue imposible ajustar un modelo integral y completo que incluyera simultáneamente constructos del estudiante y del centro educativo, con sus respectivos indicadores. Por ello fue necesario estimar

por separado un modelo estructural con variables latentes para los datos del estudiante, y un modelo de trayectoria o senderos (que solo trabaja con variables observadas) para los datos de la institución. Aun así, los resultados preliminares son muy sugestivos y proporcionan una base para ejercicios futuros más sistemáticos y ambiciosos.

Alcance y utilidad de los modelos de ecuaciones estructurales

Las raíces de los SEM se pueden trazar a partir de tres técnicas multivariadas más antiguas: el análisis de factores confirmatorio, que proviene de la Psicología, el análisis de trayectoria o de senderos (*path analysis*), que proviene de la Biología y la Genética, y los sistemas de ecuaciones simultáneas que se desarrollaron en Economía. Un modelo SEM completo integra estas tres técnicas (Kaplan, 2009; Mulaik, 2009).

Características principales

En un modelo SEM se distinguen dos componentes: el modelo estructural, que establece las relaciones causales entre los constructos, y el modelo de medición, que describe con qué indicadores son medidos los constructos (Bollen, 1989; Kaplan, 2009; Mulaik, 2009). Contar de previo con un marco teórico sólido, que sustente las hipótesis que conectan causalmente los diversos constructos involucrados en la explicación del fenómeno de interés, es esencial para el empleo de los modelos de ecuaciones estructurales, así como una correcta *operacionalización* y medición de todos los constructos y variables relevantes (recuadro 5.5).

Recuadro 5.5

Concepto de modelo

La palabra “modelo” es de uso común. Desde el punto de vista científico, tiene una acepción precisa, que es la que se utiliza en esta sección. Un modelo es una explicación teórica que describe cómo un sistema de funciones no observable, que produce predicciones de comportamiento, puede ser observado. Así, observaciones empíricas consistentes con lo predicho a partir del modelo dan evidencia a favor de su veracidad, mientras que observaciones no consistentes dan pie para refutarlo o redefinirlo.

Fuente: Elaboración propia a partir de Bollen, 1989; Freedman, 2010 y Kaplan, 2009.

Un aspecto fundamental de la técnica consiste en que las variables observadas representan un número más pequeño de constructos o variables latentes que no pueden ser directamente medidos, solo inferidos a partir de las variables observadas. El marco de referencia para la fundamentación de estos modelos plantea que hay conceptos que no se pueden medir directamente —variables latentes o constructos— pero que sí se pueden aproximar mediante indicadores que son *operacionalizaciones* de esos conceptos, y se pueden establecer relaciones de causalidad entre las variables latentes, las llamadas ecuaciones estructurales (recuadro 5.6).

En este contexto, se denomina variables exógenas a aquellas que no están determinadas por otras dentro del modelo y variables endógenas a aquellas que están explicadas por otras dentro del modelo. Usualmente se ajustan modelos lineales. Lo ideal para la estimación es que las variables sean métricas y que su distribución sea simétrica (parecida a la normal).

El uso operativo de los modelos de ecuaciones estructurales involucra al menos cuatro pasos básicos (Bollen, 1989; Kaplan, 2009):

- **Identificación:** establecer si el sistema de ecuaciones planteado tiene solución, utilizando los datos recolectados.
- **Estimación:** el método más popular sigue siendo el máximo verosímil.
- **Bondad de ajuste:** se calculan diversas medidas, basadas en el grado de similitud entre la matriz observada de variancias y covariancias, y la correspondiente matriz estimada con base en el modelo.
- **Pruebas de hipótesis:** acerca de los coeficientes del modelo.

En síntesis, con los SEM es esencial contar un claro planteamiento teórico antes de recolectar los datos y tratar de estimar empíricamente el modelo estructural. El proceso de estimación en sí mismo a menudo presenta problemas técnicos de naturaleza matemática-estadística que dificultan, o incluso pueden impedir del todo su estimación empírica. Estos problemas tienen que ver con lo que en este

Recuadro 5.6

Alcance de los modelos de ecuaciones estructurales

El hecho de que un modelo SEM sea compatible o consistente con los datos observados, es decir, que se ajuste estadísticamente a estos, solo brinda un apoyo empírico a las hipótesis de causalidad implicadas en el modelo, pero no prueba o valida el modelo causal. Se trata entonces, en lo fundamental, de una técnica de “falsación” en el sentido más clásico, pues las evidencias a favor del modelo no se prueban a partir del mismo modelo, sino eliminando posibles explicaciones alternativas. Los modelos SEM no prueban la causalidad, pero sí dan cuenta de evidencia empírica a favor o no de un planteamiento causal. Aunque siempre existirán teorías causales alternativas que explican igualmente bien las relaciones observadas, los SEM ayudan a evidenciar de manera empírica relaciones

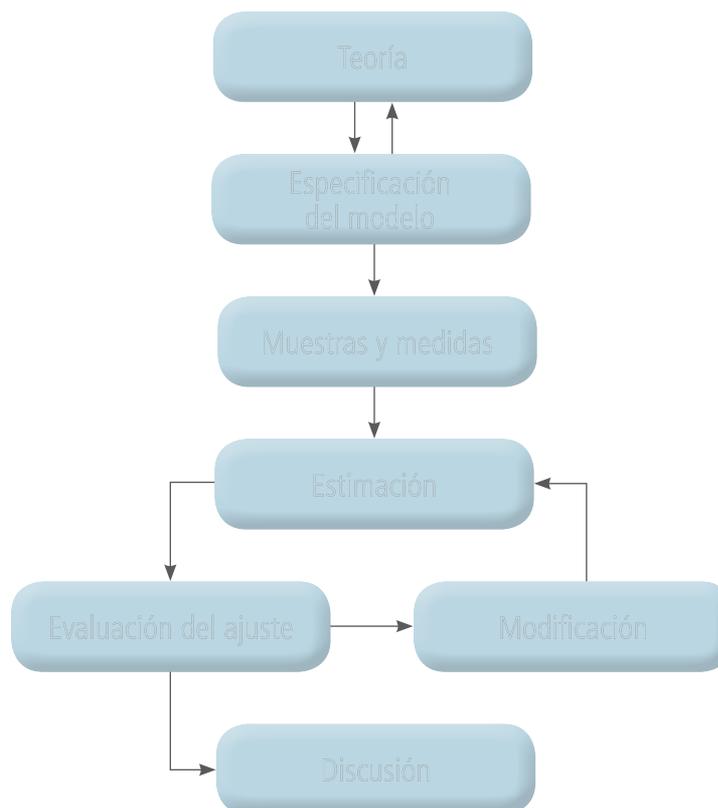
causales hipotetizadas, orientadas por la teoría y especificadas *a priori* (Bollen, 1989; Freedman, 2010; Kaplan, 2009; Ruiz et al., 2010).

Para Kaplan (2009), la secuencia de pasos involucrados en un estudio con modelos de ecuaciones estructurales se representa en la forma que muestra la figura 5.1. Las variables identificadas como relevantes se miden e incorporan a un modelo de relaciones causales hipotéticas. Entonces, los modelos estructurales son una alternativa de análisis estadístico para establecer si datos empíricos que no fueron recolectados usando un diseño experimental, favorecen un cierto planteamiento teórico causal. Otras alternativas de análisis que cumplen propósitos similares son las redes bayesianas y las redes neuronales.

Fuente: Montero, 2012.

Figura 5.1

Proceso de investigación que acompaña el uso de modelos de ecuaciones estructurales, según David Kaplan



Fuente: Kaplan, 2009.

contexto se denomina “identificación” y van desde la situación extrema en la que no es posible estimar del todo la configuración planteada, hasta aquella en la que el modelo se ajusta, pero se generan parámetros con inconsistencias. Estas últimas incluyen coeficientes estandarizados mayores a uno y el caso en que no se pueden calcular los errores estándar y, por consiguiente, tampoco las pruebas de significancia estadística individuales para cada parámetro estimado. Eventualmente se pueden resolver algunos de estos problemas aplicando estrategias que, en su mayoría, implican establecer como constantes o como nulos ciertos parámetros (Kaplan, 2009; Mulaik, 2009).

Una correcta modelación conlleva, entonces, un proceso iterativo que involucra, por un lado, una apropiada *operacionalización* de la teoría y, por otro, resolver problemas de estimación matemático-estadísticos (Freedman, 2010). De hecho, muchos modelos de análisis estadístico más conocidos y antiguos son casos particulares de un modelo SEM, entre ellos: análisis factorial (exploratorio y confirmatorio), modelos de regresión múltiple y modelos de análisis de trayectoria o de senderos. En estos últimos se trabaja únicamente con variables observadas, y se supone que están medidas sin error, por tanto no hay modelo de medición y solo se estima el modelo causal (Kaplan, 2009; Mulaik, 2009; Ruiz et al., 2010).

Ventajas con respecto a los modelos de regresión

Una de las grandes ventajas a nivel conceptual que presentan los SEM sobre los modelos de regresión (ya sea clásicos o multinivel, como los aplicados en este capítulo), es que por medio de ellos se puede generar evidencia empírica para planteamientos teóricos causales que establecen relaciones de mediación entre variables o constructos (Kaplan, 2009; Mulaik, 2009; Ruiz et al., 2010).

Para ilustrar este punto, y como se verá más adelante con los datos de la prueba PISA, la gestión del director no es un predictor directo del rendimiento en la prueba de competencia lectora, sino que su influencia sobre esta variable se da por medio de un indicador que alude a las dificultades para la enseñanza y a un clima

negativo en el colegio. De esta manera, el efecto de la gestión del director sobre el puntaje en la citada prueba está mediado por ese indicador. Se trata entonces de un efecto indirecto, que solo se presenta a través de la variable mediadora.

Modelo estructural para predecir el desempeño en competencia lectora con variables del estudiante

El modelo teórico en que se basa este ejercicio, en el caso de la prueba PISA de competencia lectora, se limita a las variables propias de los estudiantes (su nivel socioeconómico y prácticas académicas). Es decir, se trata de explicar el rendimiento en la prueba a partir de características de los individuos. Otros factores propios del centro educativo no son considerados en el ejercicio, sino que son analizados por aparte en el próximo acápite.

Teoría y variables

La teoría sobre el puntaje en la prueba PISA 2009 de competencia lectora a partir de las características individuales puede sintetizarse como sigue: el rendimiento académico depende del nivel socioeconómico de las y los estudiantes y el capital cultural de sus hogares, sus antecedentes educativos y sus prácticas académicas. Sin embargo, la influencia de estos factores no es simple, sino compleja: varios de ellos ejercen una influencia directa e indirecta sobre el rendimiento y, a su vez, están relacionados recíprocamente entre sí. La cuestión por determinar es si el modelo descarta esta proposición teórica o aporta evidencias que sugieren su relevancia.

En la especificación del modelo se establecieron tres constructos como variables exógenas: extracción social, medida por la tenencia de bienes no culturales en el hogar y el nivel educativo del padre y la madre; los antecedentes educativos, que incluye el grado que cursa el alumno y si asistió a la educación preescolar –hecho que, como lo señala el capítulo 2 de este Informe, es importante para el rendimiento académico posterior– y las estrategias de lectura, que consideran la percepción de eficacia de las estrategias para comprender y resumir textos. Son variables exógenas porque son factores que, de partida, se es-

pera determinen los puntajes que obtienen los estudiantes.

Adicionalmente, se plantearon tres constructos como variables endógenas. El primero es el capital cultural, medido por la tenencia de libros en la casa, otros bienes culturales y la frecuencia de lectura. Se espera que el capital cultural esté afectado por la extracción social de las personas (a más alta extracción, mayor capital cultural) y que, a su vez, ejerza una influencia propia, independiente, sobre el rendimiento en las pruebas. Esto por cuanto la relación entre nivel social y capital cultural no siempre es positiva.

El segundo constructo endógeno corresponde a las actitudes y hábitos de lectura. Aunque se propone que estos dependen fuertemente de las estrategias de lectura que tengan los estudiantes, se espera que los hábitos ejerzan, *per se*, una influencia propia sobre el rendimiento académico. ¿Por qué? Porque no basta con que las personas reconozcan la necesidad de aplicar estrategias para comprender y resumir textos, sino que el hecho de leer permita ejercitar esas estrategias.

Finalmente, el tercer constructo es el puntaje en la prueba PISA 2009 de competencia lectora, que es lo que se intenta explicar. Las descripciones e indicadores de los tres constructos se presentan en el cuadro 5.13.

Interpretación de resultados

El modelo estructural estimado, o modelo de medición (el sistema que conecta a los constructos con los indicadores propuestos para medir cada uno de ellos) se presenta en la figura 5.2. Casi todos los indicadores analizados presentan cargas factoriales altas, es decir, los coeficientes que representan el grado de relación entre el constructo y sus respectivos indicadores muestran valores apropiados (cargas mayores a 0,3 en valor absoluto se consideran óptimas). Solo en el caso de asistencia a la enseñanza preescolar como indicador de antecedentes educativos se presenta una relación relativamente débil (0,11). En principio, pues, no es descartable la teoría enunciada en el acápite anterior, sobre los factores explicativos del rendimiento académico en la prueba PISA 2009 de competencia lectora.

Cuadro 5.13

Constructos e indicadores utilizados en el modelo estructural, para predecir el puntaje en la prueba PISA 2009 de competencia lectora

Nombre del Constructo	Tipo	Indicadores	Errores de medida
Extracción social	Exógeno	Tenencia de bienes no directamente asociados a capital cultural	0,34
		Nivel educativo de la madre	0,66
		Nivel educativo del padre	0,66
		Clasificación de ocupación del padre	0,74
Antecedentes educativos	Exógeno	Grado que cursa	0,21
		Asistió a preescolar	0,99
		Frecuencia de repetición	0,45
Estrategias de lectura	Exógeno	Escala de percepción de la eficacia de estrategias para comprender un texto	0,80
		Escala de percepción de la eficacia de estrategias para resumir un texto	0,75
Capital cultural	Endógeno	Tenencia de bienes asociados a capital cultural	0,23
		Cantidad de libros en la casa	0,61
		Frecuencia de lectura en línea	0,59
Actitudes y hábitos de lectura	Endógeno	Tiempo de lectura por placer	0,48
		Frecuencia de lectura por iniciativa propia	0,82
		Frecuencia de lectura de textos para el colegio	0,98
		Escala de actitudes hacia la lectura	0,28
Rendimiento	Endógeno	Puntaje en la prueba de PISA de competencia lectora	0,00

Fuente: Montero, 2012.

A algún lector podría sorprenderle el signo negativo que muestra para este constructo el indicador de repetencia. Sin embargo, dada la naturaleza de esta variable, tiene sentido que se correlacione negativamente con el constructo (y por ende con los otros dos indicadores), dado que, a mayor repetencia, menor nivel en términos de antecedentes educativos.

En la dimensión de hábitos de lectura todos los indicadores presentan valores óptimos, excepto el relacionado con la lectura de textos del colegio, que tiene una relación relativamente débil (0,16). Es importante notar que el constructo “rendimiento en la prueba” solo cuenta con un indicador, que es precisamente el puntaje en la prueba PISA de competencia lectora y, por tanto, se asume que está medido sin error. Esta es una debilidad metodológica, pues lo ideal sería que cada constructo sea representado por al menos tres indicadores (Ruiz et al., 2010).

En relación con el modelo estructural, es decir, las trayectorias causales que conectan variables latentes exógenas y endógenas, se observa que los coeficientes que tienen sentido

teórico para el constructo “extracción social” son aquellos que lo conectan causalmente con “capital cultural” y con “rendimiento en la prueba” (0,90 y 0,35, en cada caso), mientras que el valor de gran magnitud, pero negativo, para la relación entre “extracción social” y “actitudes y hábitos hacia la lectura” merece estudiarse con más detenimiento, pues parece ser contra-intuitivo (figura 5.3).

La magnitud de la relación entre “antecedentes educativos” y “rendimiento en la prueba” parece razonable (0,38), mientras que la relación entre el constructo “estrategias de lectura” (comprender y resumir un texto) y “actitudes y hábitos hacia la lectura” (0,18) también resulta defendible, al igual que la relación estimada entre estas estrategias y la nota en la prueba (0,33).

Finalmente, las relaciones entre las variables endógenas son defendibles. La fuerte relación entre “capital cultural” y “actitudes y hábitos hacia la lectura” es el esperado por la teoría (0,86) —a mayor capital cultural en la casa, mejores actitudes y hábitos hacia la lectura—, al igual que la más débil relación entre “actitudes y

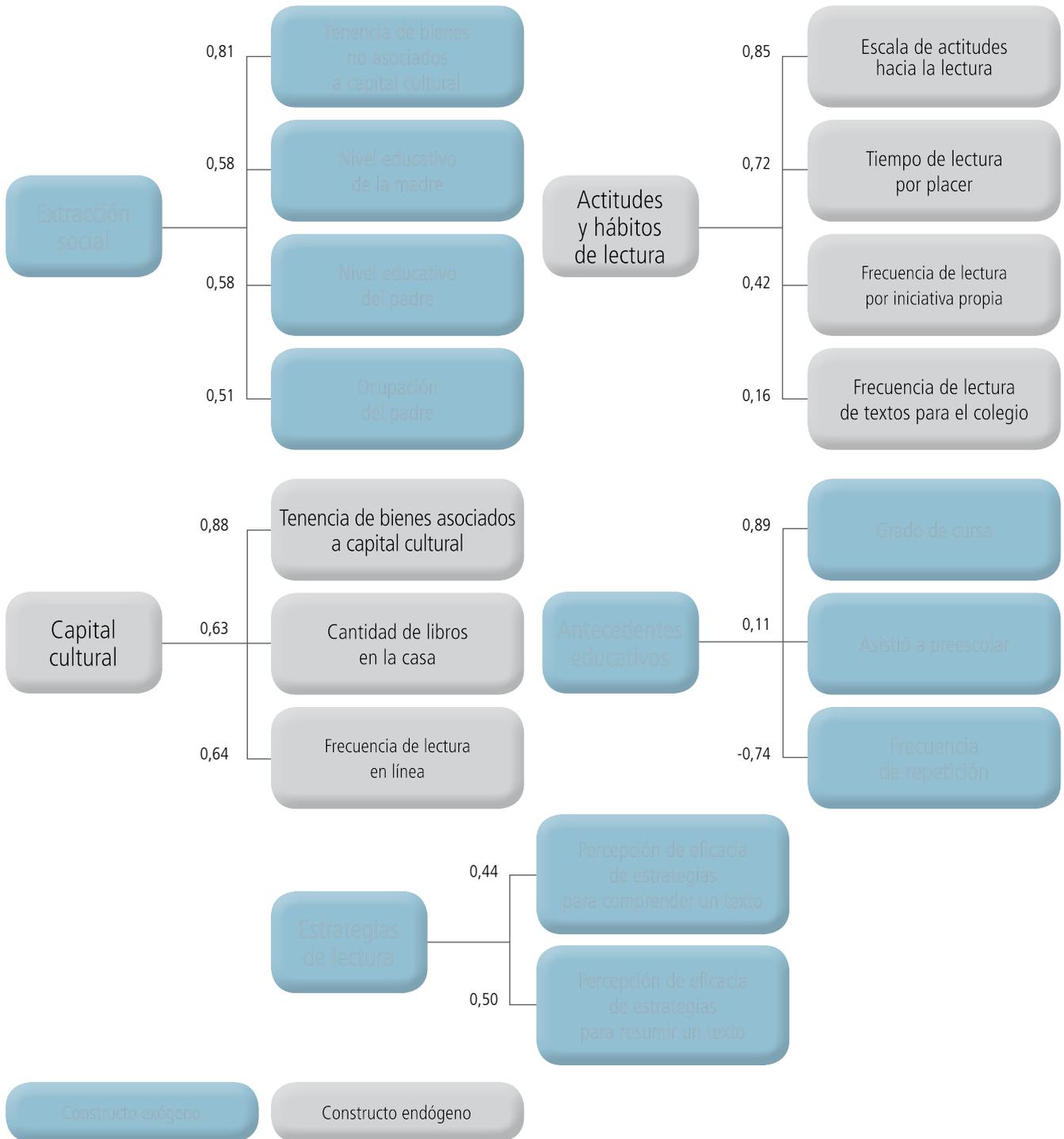
hábitos hacia la lectura” y “rendimiento en la prueba” (0,15), pues estos hábitos deberían contribuir, pero no determinar los puntajes en las pruebas estandarizadas. El capital cultural tiene así un efecto indirecto de importancia sobre el rendimiento en la prueba PISA de competencia lectora, a través de “actitudes y hábitos de lectura”, pero su efecto directo no parece tener importancia práctica (-0,08).

En cuanto a la bondad del ajuste de esta primera propuesta de modelo causal tentativo, se debe indicar que no es totalmente satisfactoria, dado que el valor óptimo ideal para la medida RMSEA (*root mean square error of approximation*)⁹ es 0,05 o menos (Kaplan, 2009; Mulaik, 2009), aunque algunos autores son más laxos y proponen el valor de 0,08 como umbral aceptable (Ruiz et al., 2010). Ello sugiere el carácter preliminar del ejercicio.

A manera de reflexión general, y en el entendido de que se trata de resultados exploratorios, que deben ser objeto de mayor análisis, es preocupante que no sean los antecedentes educativos, sino la

Figura 5.2

Indicadores seleccionados para medir los constructos en el modelo de ecuaciones estructurales para predecir el puntaje en competencia lectora, factores del estudiante^{a/}. PISA, 2009



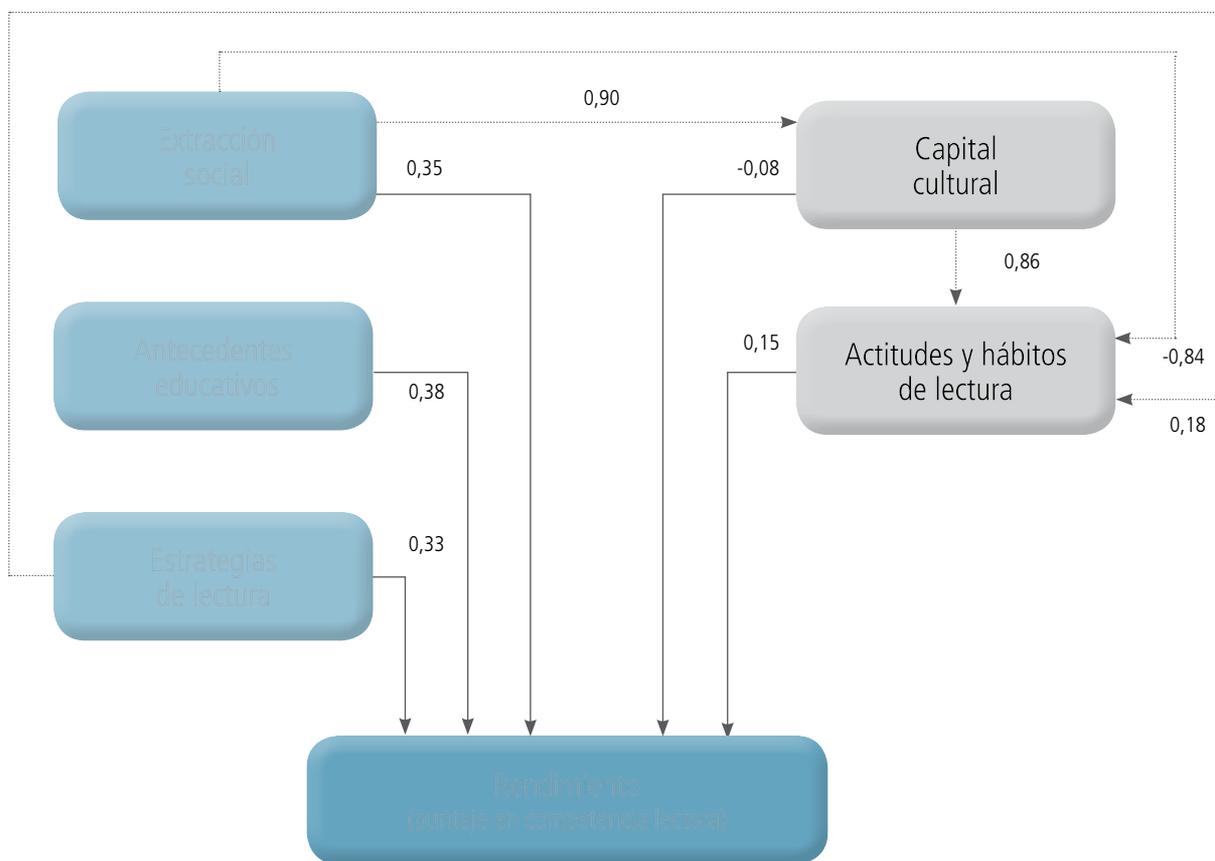
Chi-cuadrado = 1.516,1 df = 108 P = 0.000 RMSEA = 0,063

a/ Los números representan los coeficientes que conectan a cada constructo con sus respectivos indicadores, se les denomina cargas factoriales. Fluctúan en un rango de -1 a 1. Cargas mayores a 0,3 en valor absoluto son óptimas. El signo representa el tipo de relación entre el indicador y el constructo: signo positivo es una relación directa y el signo negativo es una relación inversa.

Fuente: elaboración propia con base en Montero, 2012.

Figura 5.3

Constructos del modelo de ecuaciones estructurales para predecir el puntaje en competencia lectora, factores del estudiante. PISA, 2009



—→ Efecto directo que ejerce el constructo sobre la predicción
→ Efecto indirecto que ejerce el constructo a través de otros constructos

Chi-cuadrado = 1.516,1 df = 108 P = 0.000 RMSEA = 0,063



a/ Los números representan las cargas factoriales, que fluctúan en un rango de -1 a 1. Las cargas mayores a 0,3 en valor absoluto son óptimas. El signo representa el tipo de relación: el signo positivo indica una relación directa; el negativo, inversa.

Fuente: elaboración propia con base en Montero, 2012.

extracción social del estudiante, la variable que determina con más fuerza predictiva los resultados de la prueba PISA de competencia lectora, debido, de manera conjunta, a su influencia directa (0,35) y su influencia indirecta por medio del capital cultural y las actitudes y hábitos hacia la lectura (0,90*0,86*0,15). Además, contrario a lo esperado, los antecedentes educativos no predicen de modo relevante las actitudes

y hábitos hacia la lectura. Este conjunto de resultados parece ser consistente con las brechas en las oportunidades educativas que las crecientes desigualdades sociales y económicas están generando en el país.

Los hallazgos también ofrecen evidencia de posibles deficiencias en la aplicación de los programas educativos en una dimensión de impacto fundamental: las actitudes y hábitos hacia la lectura. Desde una óptica

de lo que “debería ser”, en una sociedad menos segregada en lo social y lo económico, la extracción social del estudiante debería tener un peso relativamente menor en el rendimiento en pruebas estandarizadas y en otras variables de impacto educativo. Por el contrario, los antecedentes educativos de los alumnos (algo que se supone más fácil de modificar con acciones de política) deberían tener una influencia mayor. Por

otra parte, la predicción tan precisa que se puede hacer en la dimensión de actitudes y hábitos hacia la lectura a partir del capital cultural, preocupa también desde la perspectiva de la equidad. La relación directa del capital cultural en el puntaje de PISA no es de relevancia (-0,08), sino que se presenta a través de las actitudes y hábitos de lectura.

Por último, cabe destacar la influencia de las estrategias de lectura (para comprender y para resumir textos) tanto en las actitudes y hábitos hacia la lectura, como en el mismo puntaje de la prueba (coeficientes de 0,18 y 0,33, respectivamente). A partir de estos resultados la recomendación evidente al sistema educativo es promover la enseñanza explícita de estas estrategias.

Modelo estructural para predecir el desempeño en competencia lectora con las variables del centro educativo

En el caso de los factores institucionales (variables del director y del centro educativo) solamente fue posible ajustar un modelo de análisis de trayectoria o de senderos, en el que no existe modelo de medición y se asume que las variables se miden sin error. Esto se debe a problemas de estimación matemático-estadísticos, muy comunes en el contexto de los SEM.

Teoría y variables

Es importante señalar que la disponibilidad de variables para estudiar los factores institucionales fue muy poca. Solo fue posible incluir dos constructos, en calidad de variables exógenas. En primer lugar, la gestión del director, que se mide con una escala y, en segundo lugar, el índice de desarrollo social (IDS) del distrito en que se ubica el colegio. Se utilizaron como referencia los datos del IDS para el año 2007, los valores con desagregación distrital más cercanos a la fecha en que se realizó la prueba.

Adicionalmente, hay cuatro variables endógenas: i) un índice de dificultades para la enseñanza y clima escolar negativo, ii) el perfil del colegio, que es una escala ordinal compuesta por la combinación de las variables dependencia (público o privado), modalidad (académico o técnico) y zona (urbano o rural). El perfil del colegio toma el valor más bajo en el caso

de público rural académico y el más alto en el caso de privado urbano académico, iii) el promedio de computadoras por estudiante que presenta el colegio y iv) el puntaje en la prueba PISA de competencia lectora.

En términos generales, la teoría es que el puntaje de las y los estudiantes en la prueba PISA 2009 de competencia lectora es afectado directamente por el desarrollo social del entorno y la gestión del director del centro educativo. Estos factores también inciden de manera indirecta en el rendimiento académico, según el tipo de colegio y el clima escolar medido a partir de las computadoras disponibles por estudiante, variables que a su vez tienen una influencia propia.

Interpretación de resultados

Los resultados del modelo de ecuaciones estructurales con variables del entorno se presentan en la figura 5.4. Aparte de las dificultades técnicas ya mencionadas para estimar un modelo SEM completo, el modelo de trayectoria para los factores institucionales dista de presentar niveles óptimos de bondad de ajuste. Por tanto, los resultados deben considerarse preliminares y tentativos. Debe indicarse, eso sí, que todos los coeficientes tienen el signo esperado de acuerdo con la teoría y esto permite hacer una interpretación más directa de los hallazgos, aunque los comentarios deben tomarse con cautela por el carácter preliminar del ejercicio.

Desde el punto de vista de la equidad, resulta preocupante que el IDS figure como una variable predictora relevante para tres de las cuatro variables endógenas (índice de dificultades para la enseñanza y clima escolar negativo, perfil del colegio y puntaje en competencia lectora). Esta es quizá la forma más injusta de inequidad, pues relaciona variables educativas individuales del estudiante con el nivel de desarrollo del distrito en que se ubica su colegio; en ambos casos se trata de desigualdades de tipo estructural, que deben atenderse también estructuralmente.

Por otra parte, como ya se indicó, la gestión del director o directora tiene un efecto indirecto en los puntajes de la prueba, a través del índice de dificultades para la enseñanza y clima escolar negativo. Entonces, la recomendación de política

educativa es mejorar las destrezas de gestión del director, a fin de atenuar los problemas para la enseñanza y el clima negativo, lo que a su vez incidirá en mejores desempeños en PISA.

Resulta interesante discutir la relación directa encontrada entre el número de computadoras por estudiante y el puntaje de la prueba de competencia lectora. Quizá se trate de un resultado espurio, debido al efecto de otra variable asociada a estas dos pero no controlada: el acceso a material de lectura que permite una computadora conectada a Internet. Parece pertinente, eso sí, recomendar, como medida de política educativa que se aumente el número de computadoras por estudiante en los colegios, como una forma de coadyuvar a que los estudiantes tengan mayor oportunidad de realizar lecturas en línea.

Por último, se reitera que la situación ideal a nivel de análisis sería integrar simultáneamente, en un solo “megamodelo” estructural, variables del estudiante y de la institución. Cuando se estima cada uno por separado —como se hizo en este apartado— se pueden generar sesgos de estimación en los parámetros, situación que se denomina “falacia atomística” si el modelo solo utiliza variables del estudiante, y “falacia ecológica” si solo se incluyen variables de la institución (Gaviria y Castro, 2005).

Principales desafíos en materia de rendimiento académico

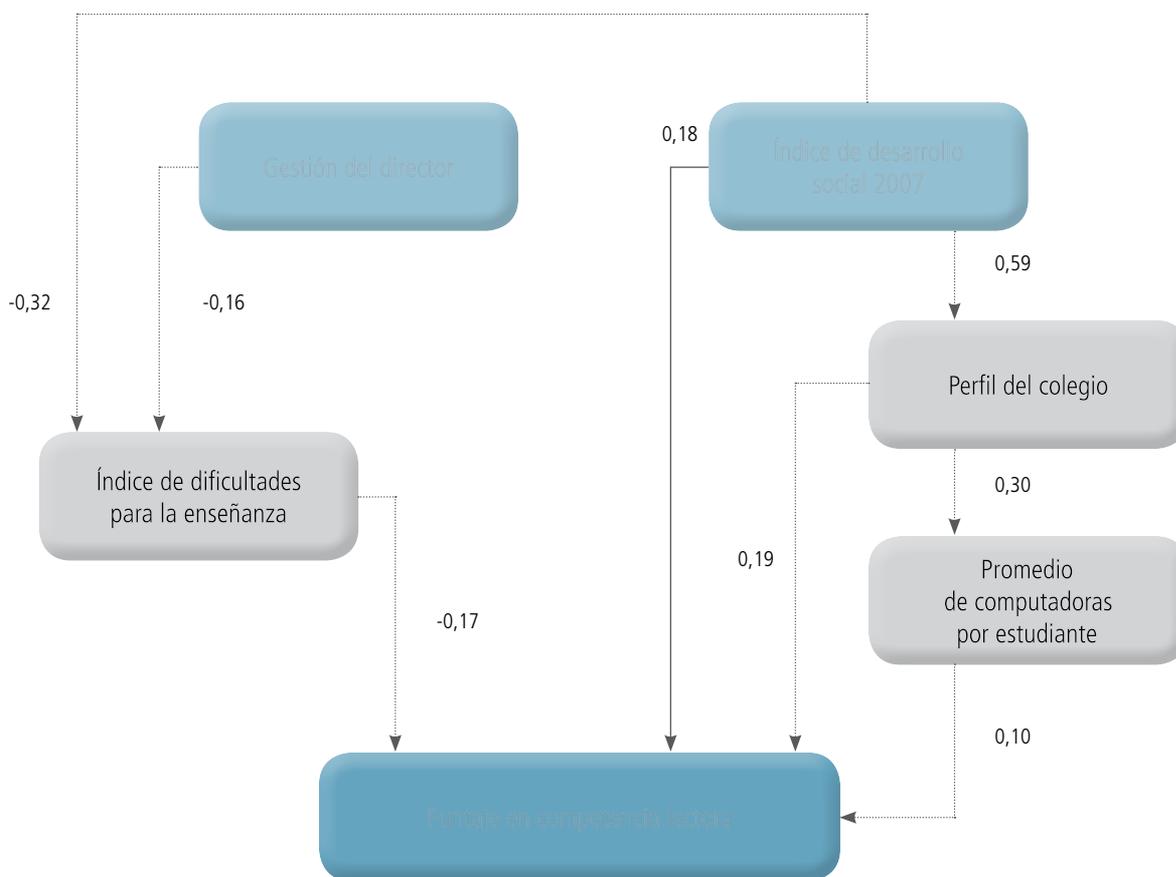
La exploración de los factores asociados al rendimiento académico mediante la aplicación de modelos de análisis multinivel, ha permitido visualizar una serie de desafíos para elevar el desempeño de los estudiantes costarricenses en pruebas estandarizadas. Estos desafíos se desprenden de la evidencia recabada y su identificación y posteriores reflexiones cierran el presente capítulo especial.

Con el propósito de mejorar los niveles de rendimiento en la prueba PISA de competencia lectora, es recomendable poner en marcha las siguientes acciones:

- Promover en los estudiantes una actitud positiva hacia la lectura.
- Enseñar explícitamente en el aula las estrategias que son eficaces para resumir un texto.

Figura 5.4

Modelo de trayectoria para predecir el puntaje de competencia lectora, factores institucionales^{a/}. PISA, 2009



—> Efecto directo de la variable exógena sobre la predicción

⋯> Efecto indirecto entre variables endógenas

Chi-cuadrado = 135,0 df = 4 P = 0.000 RMSEA = 0,085

a/ Los números representan las cargas factoriales, que fluctúan en un rango de -1 a 1. Cargas mayores a 0,3 en valor absoluto son óptimas. El signo representa el tipo de relación: signo positivo es una relación directa y el signo negativo es una relación inversa.

Fuente: Elaboración propia con base en Montero, 2012.

- Enseñar explícitamente en el aula las estrategias que son eficaces para comprender un texto.
- Promover la lectura en línea.
- Promover la lectura por iniciativa propia.
- Aumentar el número de lecciones de Español en aquellos colegios donde los estudiantes reportan cantidades inferiores al promedio.
- 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.
- Promover la lectura en línea.
- Promover en los estudiantes una actitud positiva hacia la lectura.
- Aumentar el número de lecciones de Matemática en aquellos colegios donde los estudiantes reportan cantidades inferiores al promedio.
- Promover la lectura por iniciativa propia.
- Enseñar en el aula estrategias apropiadas para el uso de técnicas analíticas al estudiar.
- 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 falta de pertinencia del material pedagógico, entre otros.

Por otra parte, el hecho de que el IDS del distrito en que se ubica el colegio resulte

relevante como predictor del rendimiento académico llama a la reflexión en cuanto a la equidad en términos geográficos y de políticas de desarrollo, pues por la sola razón de residir en un distrito con bajo IDS es más probable que un estudiante obtenga puntajes relativamente bajos en las pruebas PISA. Esta circunstancia sin duda está asociada a las diferencias de acceso a las oportunidades para aprender, que probablemente se presentan según zonas geográficas, y merece atención por parte de investigadores y tomadores de decisiones. La influencia de la condición socioeconómica también es determinante en el caso de las pruebas diagnósticas del MEP. En virtud de que este es un factor dado y que los estudiantes pueden hacer poco para cambiarlo, se enfatiza en la necesidad de formular acciones para aminorar los efectos de las diferencias en las brechas socioeconómicas.

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 (OCDE, 2012a). La relevancia de la lectura y su contracara, la escritura, sumada al hecho de que ambas destrezas no se desarrollan naturalmente, sino que deben adquirirse por medio de la enseñanza sistemática, son una razón de la existencia de las pruebas PISA. Así lo explicitan sus promotores y diseñadores (OCDE, 2012a):

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.

Para llamar aun más la atención sobre la importancia de dirigir acciones de política educativa hacia el desarrollo de las dimensiones de la lectura, se ejecutó un estudio de simulación en el que 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 de los alumnos de colegios privados en las ya citadas variables de lectura. Las diferencias observadas se atribuirían a la mejora de las competencias lectoras en los colegios públicos (cuadro 5.14).

Esta simulación sugiere que el mejoramiento de las competencias de lectura en los colegios públicos disminuiría de modo sustancial las brechas existentes entre alumnos de colegios públicos y privados en el desempeño en las pruebas PISA. Con respecto a las diferencias actuales, la reducción sería de más de 67% en competencia lectora y superior al 70% en competencia matemática. Por lo tanto, se recomienda ejecutar acciones de política educativa dirigidas a promover asertivamente las dimensiones de la lectura en los colegios públicos.

Otro de los desafíos se relaciona con el bajo nivel de representatividad de las pruebas PISA, que tan solo cubren al 53% de la población de 15 años. Este hecho obliga al país a asumir un compromiso para atender las necesidades educativas de ese otro 47% de jóvenes que no están representados en estos análisis y que son, probablemente, los más vulnerables y desaventajados. También sería conveniente aplicar las pruebas estandarizadas a los estudiantes que tienen adecuaciones curriculares, con el propósito de identificar problemas y limitaciones en sus procesos de aprendizaje.

Finalmente, cabe reiterar la importancia de la motivación y las expectativas familiares con respecto al desempeño estudiantil. Los conocimientos y destrezas que las y los estudiantes adquieren en los centros educativos son herramientas esenciales, pero insuficientes, para su éxito académico. El apoyo del núcleo familiar es sin duda otro componente clave para el logro de ese objetivo.

Cuadro 5.14

Simulación del desempeño de estudiantes de colegios públicos en las pruebas PISA, 2009^{a/}

Construc ^{to}	Simulación			Valores reales		
	Públicos	Privados	Diferencia	Públicos	Privados	Diferencia
Competencia lectora	467,0	493,0	25,9	398,6	478,3	-79,6
Competencia matemática	410,4	435,8	25,4	432,0	519,7	-87,7

a/ Si los estudiantes tuvieran los niveles promedio en las variables de lectura que poseen los alumnos de colegios privados.

Fuente: Montero et al., 2012.

Anexo

Cuadro A.1

Niveles de desempeño en la escala de competencia lectora en las pruebas PISA

Nivel de desempeño	Límite inferior de puntaje	Características de las tareas
6	698	Las tareas de este nivel demandan 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 soliciten al lector que 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.
5	626	Las tareas de recuperación de la información en este nivel requieren que el lector localice y organice varias piezas de información poco prominente, y deduzca cuáles datos del texto son relevantes. Las tareas de reflexión demandan una evaluación o hipótesis crítica, basada en conocimientos especializados. Tanto las tareas de interpretación como las de reflexión requieren una comprensión completa y detallada de un texto cuyo contenido o formato es poco conocido. En todos los aspectos de lectura, las tareas de este nivel normalmente implican tratar con conceptos contrarios a lo esperado.
4	553	Para realizar las tareas de recuperación de la información en este nivel el lector debe localizar y organizar varias piezas de información poco prominente. Algunas tareas demandan una interpretación del significado de los matices del lenguaje en una sección del texto, teniendo en cuenta el documento en su conjunto. Otras tareas de interpretación pueden requerir la comprensión y aplicación de categorías dentro de un contexto poco conocido. Las tareas de reflexión solicitan que los lectores utilicen conocimientos formales o públicos para establecer hipótesis o evaluar críticamente un texto. Los lectores deben demostrar una comprensión precisa de textos largos o complejos, cuyo contenido o formato puede ser poco familiar.
3	480	Las tareas de este nivel requieren que el lector localice, y en algunos casos reconozca, la relación entre varias piezas de información que deben cumplir múltiples condiciones. En las tareas de interpretación se solicita que el lector integre varias partes del texto para identificar una idea principal, comprender la relación o llegar al significado de una palabra o frase. Debe tener en cuenta muchos rasgos para comparar, contrastar y categorizar. A menudo la información necesaria no es muy prominente o hay muchos datos en conflicto, o bien existen otros obstáculos en el texto, como ideas contrarias a lo esperado o redactadas en negativo. Las tareas de reflexión de este nivel pueden implicar conexiones, comparaciones y explicaciones, o pueden exigir que el lector evalúe una característica del texto. Algunas tareas de reflexión requieren que el lector demuestre una fina comprensión del texto en relación con conocimientos familiares y cotidianos. Otras tareas no demandan una comprensión detallada del texto, pero sí que el lector recurra a conocimientos menos comunes.
2	407	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 piden 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 pueden comprender comparaciones o contrastes basados en una única característica del texto. Las típicas tareas de reflexión de este nivel solicitan que los lectores realicen comparaciones o varias conexiones entre el texto y conocimientos externos, recurriendo a experiencias y actitudes personales.
1a	335	Las tareas de este nivel requieren que el lector localice una o más piezas independientes de información explícitamente indicada; que reconozca el tema principal o el propósito del autor en un texto sobre un tema conocido, o realice una sencilla conexión entre información en el texto y conocimientos comunes cotidianos. Por lo general la información requerida del texto es prominente y hay poca, o ninguna, información en conflicto. Al lector se le dirige explícitamente a considerar factores relevantes en la tarea y en el texto.
1b	262	Las tareas de este nivel demandan que el lector localice una única pieza de información explícita en una posición prominente en un texto breve, sintácticamente sencillo, con un contexto y tipo de texto conocidos, como una narración o listado sencillos. Por lo general el texto ofrece ayuda al lector, mediante la repetición de la información, imágenes o símbolos conocidos. La información en conflicto es mínima. En las tareas que requieren interpretación, es posible que el lector tenga que realizar sencillas conexiones entre piezas de información adyacentes.

Cuadro A.2**Niveles de desempeño en la escala de competencia matemática en las pruebas PISA**

Nivel	Límite inferior de puntaje	Características de las tareas
6	669	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. Puede relacionar la información proveniente de diversas fuentes o representaciones. Además debe ser capaz de realizar razonamientos matemáticos avanzados, 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, e interpreta y reflexiona apropiadamente sobre sus hallazgos.
5	607	En este nivel el alumno puede desarrollar y trabajar con modelos de situaciones complejas, así como identificar las limitaciones de los mismos y especificar los supuestos considerados. Puede seleccionar, comparar y evaluar estrategias adecuadas para la resolución de problemas y aplicarlas para resolver situaciones complejas relacionadas con los modelos que se plantea. Asimismo, el estudiante de este nivel puede seleccionar estrategias apropiadas y utilizar en forma adecuada sus habilidades de pensamiento y razonamiento. Establece relaciones entre las representaciones, la simbología y las caracterizaciones formales. Por último, es capaz de reflexionar sobre sus decisiones y está en capacidad de formular y comunicar sus interpretaciones y los razonamientos efectuados para obtener sus respuestas.
4	545	En este nivel el estudiante trabaja eficazmente con modelos explícitos de situaciones complejas concretas, las cuales pueden implicar ciertas restricciones o suposiciones. Es capaz de seleccionar e integrar diferentes representaciones, incluyendo representaciones simbólicas, y vincularlas directamente con situaciones del mundo real. Además puede utilizar las habilidades desarrolladas previamente y es capaz de razonar en forma competente. El alumno puede elaborar y comunicar explicaciones bien fundadas y basadas en las interpretaciones de las acciones realizadas para resolver un problema determinado.
3	482	En este nivel el estudiante puede ejecutar procedimientos claramente descritos, incluyendo aquellos que requieren la toma de decisiones secuenciales. Es capaz de seleccionar y aplicar estrategias sencillas de resolución de problemas. También puede interpretar y utilizar representaciones tomadas de diversas fuentes de información y razonar directamente sobre ellas. Puede elaborar un reporte breve para comunicar las interpretaciones y razonamientos realizados y los resultados obtenidos.
2	420	En este nivel el alumno puede interpretar y reconocer situaciones en contextos que requieren únicamente de la inferencia directa. Puede extraer la información relevante a partir de una sola fuente y utilizar un único modo de representación. Además puede emplear algoritmos básicos, fórmulas y procedimientos convencionales. También realiza razonamientos directos e interpretaciones literales de los resultados.
1	358	El estudiante de este nivel puede responder preguntas que involucran contextos familiares, que están claramente definidas y en las que toda la información relevante está presente. Es capaz de identificar la información necesaria para llevar a cabo procedimientos rutinarios siguiendo instrucciones directas en situaciones explícitas. Puede realizar acciones obvias y continuar de inmediato con determinados estímulos.
Debajo de nivel 1	Menor de 358	En este nivel el alumno no tiene éxito en las tareas matemáticas más básicas medidas por PISA. Su patrón de respuestas hace prever que no sería capaz de resolver la mitad de las tareas en una prueba elaborada exclusivamente con preguntas del nivel 1. Este estudiante tiene serias dificultades para utilizar la Matemática en cualquier contexto.

Fuente: Salas, 2012.

La coordinación del capítulo y la elaboración del documento base estuvieron a cargo de Eiliana Montero, Shirley Rojas, Ana María Rodino y Evelyn Zamora participaron como investigadoras asociadas.

La edición final fue realizada por Ronald Alfaro, con el apoyo de Jorge Vargas Cullell y Jennyfer León.

Se elaboraron los siguientes insumos: “Constructo de alfabetización matemática según PISA”, de Óscar Salas; “La competencia lectora de los estudiantes costarricenses según la evaluación internacional PISA 2009+”, de Ana María Rodino; “Costa Rica en las pruebas PISA 2009 de competencia lectora y alfabetización matemática”, de Eiliana Montero, Shirley Rojas, Evelyn Zamora y Ana María Rodino; “Costa Rica en PISA: experiencias”, de Pablo Mena; “Costa Rica: pruebas PISA. Constructos teóricos: español y matemáticas”, de Eiliana Montero; “El modelo de Rasch: una herramienta esencial para la evaluación educativa de gran escala”, de Shirley Rojas; “Reforma de la educación matemática en Costa Rica: avances y desafíos”, de Ángel Ruiz. Además se contó con dos contribuciones especiales, ambas publicadas por el MEP: “Síntesis de los factores asociados al rendimiento académico de los estudiantes de noveno año en la prueba diagnóstica de Español de tercer ciclo de la educación general básica”, de Olga Muñoz, y “Síntesis de los factores asociados al rendimiento académico en Matemática en el tercer ciclo de la educación general básica: un estudio multinivel”, de Yessenia Oviedo.

El procesamiento de las bases de datos estuvo a cargo de Shirley Rojas, quien además preparó un análisis sobre la calidad de los ítems empleados para evaluar el desempeño en Matemática.

Por sus aportes, contribuciones especiales y comentarios se agradece a: Pablo Mena, Ángel Ruiz, Juan Manuel Esquivel, Shirley Rojas y Ana María Rodino. Igualmente se agradece a Félix Barrantes y Lilliam Mora, de la Dirección de Gestión y Evaluación de la Calidad del MEP, por facilitar el acceso a la información y a Maurice Walker por la atención de las consultas realizadas.

El taller de consulta se realizó el 16 de octubre, con la participación de: Karol Acón, Ángel Alvarado, Eduardo Barahona, Jocklin Barrantes, Antonio Briceño, Víctor Buján, Carlos Castro, Karen Chacón, Rigoberto Corrales, Juan Manuel Esquivel, Ronny Gamboa, Maleni Granados, Miguel Gutiérrez, Fabio Hernández, María Jiménez, Jennyfer León, Roxana Martínez, José Andrés Masis, Eiliana Montero, Lilliam Mora, Pablo Mora, Tania Moreira, Leda Muñoz, Olga Muñoz, Dagoberto Murillo, Yerry Murillo, Jessenia Oviedo, Ana María Rodino, Shirley Rojas, Isabel Román, Ángel Ruiz, Oscar Salas, Irene Salazar, Irene Satherland, Rocío Torres, Jorge Vargas, Carlos Villalobos y Renata Villers.

La revisión y corrección de cifras fue realizada por Jennyfer León.

Notas

- Uno de los temas más sensibles en esta materia es el de la equidad. Por ello estas disciplinas cuentan con herramientas técnicas para evitar los sesgos derivados del origen o *background* de los examinados y para que las pequeñas diferencias de dificultad en distintas versiones de la prueba sean eliminadas y no afecten la comparabilidad de las puntuaciones (procedimientos de *equating* o equiparación de puntajes).
- Es interesante observar los cambios que ha tenido la escala de competencias a lo largo de la existencia de estas pruebas. En PISA 2000 se trabajó con cinco niveles de lectura, de los cuales el 5 era el de mayor complejidad y el 1 el más elemental. En el ciclo 2009 los niveles extremos se desagregaron a fin de obtener información más precisa: se agregó un nivel 6, que refleja un desempeño altamente destacado (una competencia lectora muy diestra y diversificada) y el nivel 1 se desagregó en 1a y 1b para distinguir mejor las capacidades de lectura de los alumnos de peor desempeño.
- Las referencias que aparecen antecedidas por la letra “E” corresponden a entrevistas o comunicaciones personales realizadas durante el proceso de elaboración de este informe. La información respectiva se presenta en la sección de “Entrevistas”, de las referencias bibliográficas de este capítulo.
- Los principales insumos para este estudio fueron las bases de datos de las pruebas PISA 2009, disponibles en el sitio *web* de esa organización. De estas se seleccionaron los registros correspondientes a Costa Rica. Luego se unieron las bases de datos de los cuestionarios aplicados a estudiantes y directores, de las cuales se tomó la información relevante para el caso costarricense. Se procedió entonces a realizar las transformaciones y recodificaciones necesarias para que los datos pudieran introducirse válidamente en el modelo multinivel. Como paso previo al ajuste del modelo, se eliminaron las variables con porcentajes relativamente altos de valores faltantes y las que presentaron muy escasa variabilidad. Los análisis se realizaron con los paquetes SPSS y Stata. En este último se utilizó un algoritmo especializado para efectuar una regresión multinivel con los datos de PISA. Los modelos empleados tienen dos niveles de agregación: en el primero se incluyen características de los estudiantes que realizaron las pruebas y en el segundo se incorporan aspectos propios de los colegios a los que asisten estos jóvenes.
- El porcentaje debido a factores institucionales puede deberse a que el tamaño de la muestra de colegios para el modelo multinivel no es muy elevado ($n=163$). En la muestra de estudiantes los valores faltantes en muchas de las variables provocan que la muestra efectiva para el análisis multinivel se reduzca a 3.203 observaciones.
- Este es el caso de la variable “lecciones fuera del horario escolar”, cuyo coeficiente 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, cuanto mayor es el número de lecciones extra que recibe el estudiante, menor será su puntaje en PISA. Una posible explicación de este comporta-

miento espurio es que los alumnos de menor capacidad intelectual son los que tienden a requerir este tipo de apoyo. Este resultado también llama a la reflexión con respecto a la naturaleza y eficacia de las 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. Para efectos de este capítulo se considera como un resultado espurio y no refleja una

relación de causalidad. Situaciones análogas se dieron con las variables “número de estudiantes en la clase de Español” y “frecuencia de realización de tareas de lectura”.

- 7 Como se verá más adelante, la baja capacidad predictora de las variables de nivel 2 y 3 se debe a que la agrupación de individuos por docente en el nivel 2, y por director en el nivel 3, reduce considerablemente la variabilidad de los puntajes de la prueba y disminuye la bondad de ajuste de los modelos multinivel. En este caso particular, lo ideal hubiese sido que los individuos se
- 8 Las SEM también se denominan modelos de LISREL (*linear structural relations*), que es el nombre del paquete de *software* más antiguo especializado en su estimación, modelación o modelamiento causal y estructuras de covariancia.
- 9 Es un mecanismo de ajuste suplementario que se usa para muestras grandes en las que se utiliza el Chi-cuadrado.

agruparan por aula o sección (nivel 2) y por centro educativo (nivel 3). Sin embargo, no se dispuso de los datos desagregados para analizar la variabilidad por aula y centro educativo.

