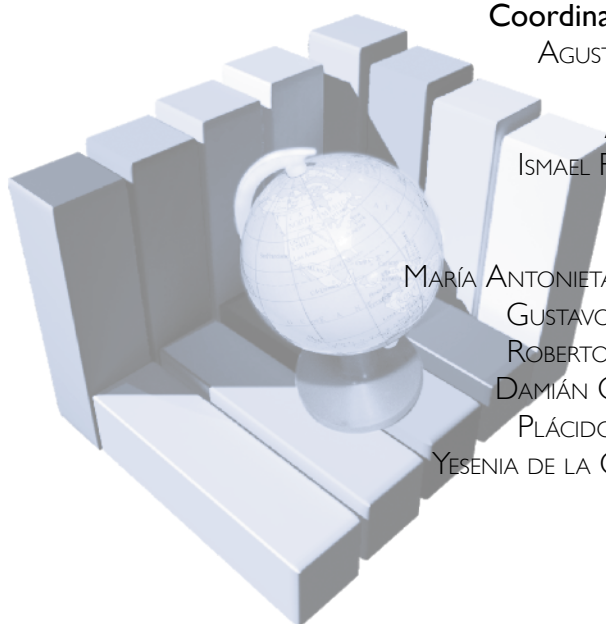


ANÁLISIS MULTINIVEL DE LA CALIDAD EDUCATIVA EN MÉXICO ANTE LOS DATOS DE PISA 2006



Coordinador del Estudio

AGUSTÍN TRISTÁN LÓPEZ

Asesor Externo

ISMAEL FLORES-CERVANTES

Por el INEE:

MARÍA ANTONIETA DÍAZ GUTIÉRREZ

GUSTAVO FLORES VÁZQUEZ

ROBERTO SOLÍS GONZÁLEZ

DAMIÁN CANALES SÁNCHEZ

PLÁCIDO MORELOS MORA

YESENIA DE LA CRUZ HERNÁNDEZ

Diciembre 2008



ANÁLISIS MULTINIVEL DE LA CALIDAD EDUCATIVA EN MÉXICO ANTE LOS DATOS DE PISA 2006

Coordinación editorial:

Miguel A. Aguilar R.
Silvia A. Tapia C.

Diseño y formación:

Luis E. Ramírez Juárez

INSTITUTO NACIONAL PARA LA EVALUACIÓN DE LA EDUCACIÓN

José Ma. Velasco 101- 5°. Piso, Col. San José Insurgentes,
Delegación Benito Juárez, México, 03900, D.F.

Primera edición 2008

El contenido, la presentación, así como la disposición en conjunto y de cada página de esta obra son propiedad del editor. Se autoriza su reproducción parcial o total por cualquier sistema mecánico, electrónico y otros, citando la fuente.

Impreso en México

ISBN: en trámite.

Contenido

Resumen	7
Presentación	11
Capítulo 1. El Modelo de Calidad Educativa del INEE	15
1.1 Necesidad de un Modelo de Calidad de la educación	17
1.2 Modelo de Calidad Educativa del INEE	19
Capítulo 2. Metodología del estudio	23
2.1 Descripción general del proyecto PISA	25
2.1.1 Presentación	25
2.1.2 Los ciclos	25
2.1.3 Población objetivo	26
2.1.4 Instrumentos	26
2.1.5 Muestra	27
2.1.6 Países participantes	27
2.2 Descripción de la metodología para el estudio	27
2.2.1 Escenarios para el desarrollo de las hipótesis	28
2.2.2 Formulación de hipótesis	29
2.2.3 Metodología para el estudio	30
Capítulo 3. Análisis	33
3.1 Preliminares	35
3.1.1 Análisis realizados	35
3.1.2 Análisis de validez de la escala de la prueba PISA 2006	36
3.1.3 Agrupación de variables	39

3.2 Modelo incondicional	42
3.3. Modelo de la situación del estudiante	44
3.3.1 Pregunta de investigación	44
3.3.2 Hipótesis	45
3.3.3 Variables explicativas	45
3.3.4 Resultados del análisis factorial confirmatorio	45
3.3.5 Ecuación del modelo multinivel	46
3.3.6 Presentación e interpretación de resultados	48
3.3.7 Conclusiones e implicaciones de los resultados	50
3.4 Modelos de la dimensión de equidad	52
3.4.1 Pregunta de investigación	52
3.4.2 Hipótesis	53
3.4.3 Variables explicativas	53
3.4.4 Resultados del análisis factorial confirmatorio	54
3.4.5 Ecuaciones de los modelos multinivel	57
3.4.6 Presentación e interpretación de resultados	59
3.4.7 Conclusiones e implicaciones de los resultados	64
3.5 Modelos para la dimensión de eficiencia	68
3.5.1 Pregunta de investigación	68
3.5.2 Hipótesis	68
3.5.3 Variables explicativas	68
3.5.4 Resultados del análisis factorial confirmatorio	69
3.5.5 Ecuaciones del modelo multinivel	72
3.5.6 Presentación e interpretación de resultados	74
3.5.7 Conclusiones e implicaciones de los resultados	75
3.6 Modelos para la dimensión de impacto	76
3.6.1 Pregunta de investigación	76
3.6.2 Hipótesis	77
3.6.3 Variables explicativas	77
3.6.4 Resultados del análisis factorial confirmatorio	77
3.6.5 Ecuaciones del modelo multinivel	79
3.6.6 Presentación e interpretación de resultados	81
3.6.7 Conclusiones e implicaciones de los resultados	83
3.7 Modelo de la dimensión de eficacia	84
3.7.1 Pregunta de investigación	84
3.7.2 Hipótesis	84
3.7.3 Variables explicativas	85
3.7.4 Resultados del análisis factorial confirmatorio	85
3.7.5 Ecuaciones del modelo multinivel	88

3.7.6 Interpretación de resultados	89
3.7.7 Conclusiones e implicaciones de los resultados	92
3.8 Modelo de la dimensión de pertinencia y relevancia	93
3.8.1 Pregunta de investigación	93
3.8.2 Hipótesis	94
3.8.3 Variables explicativas	94
3.8.4 Resultados del análisis factorial confirmatorio	94
3.8.5 Ecuaciones del modelo multinivel	96
3.8.6 Presentación e interpretación de resultados	97
3.8.7 Conclusiones e implicaciones de los resultados	98
Conclusiones	101
Referencias	111
Anexos	119
Anexo 1. La calidad como variable global	119
Anexo 2. Especificaciones para los modelos del estudio	121
A2.1 Variables	132
A2.1.1 Descripción de las variables disponibles en las bases de datos de PISA y otros datos tomados de estudios propios del INEE	134
A2.1.2 Comentarios sobre las variables empleadas en el estudio	143
A2.1.2.1 Índice de nivel económico, social y cultural	143
A2.1.2.2 Rezago académico del estudiante	144
A2.1.2.3 Regionalización de México	145
A2.1.2.4 Marginación	147
A2.1.2.5 Ubicación decilar del estudiante	147
A2.1.2.6 Variables no cognitivas	148
A2.1.3 Sobre el funcionamiento diferencial por género	148
A2.2 Imputación de datos faltantes	149
A2.2.1 Sobre las omisiones en la recopilación de datos del proyecto PISA 2006	149
A2.2.2 Imputación en casos faltantes	149
A2.2.3 Comentarios sobre los métodos de tratar las omisiones	150
A2.2.4 La imputación en la base de datos de PISA	152
A2.2.5 La imputación relacionada con los programas de análisis multinivel y de ecuaciones estructurales	153
A2.2.6 Imputación para los datos de omisión en este estudio	153

A2.3 Sobre los pesos o ponderaciones individuales y regionales	155
A2.3.1 Cálculo de las ponderaciones de los estudiantes	156
A2.3.2 Cálculo de los índices de regiones	156
A2.4 Sobre los modelos multivariados de ecuaciones estructurales	157
A2.5 Modelos multinivel	161
A2.6 Sobre el uso de variables <i>bandera</i> en los modelos multinivel	165
A2.7 Respecto de los valores plausibles	170
A2.8 Criterios para aceptar o rechazar las hipótesis	174
A2.8.1 Criterios de aceptación del modelo	174
A2.9 Programas utilizados en el estudio	177
A2.10 Archivos de datos	178
Anexo 3. Análisis multinivel de los modelos incondicionales	179
A3.1 Primer caso de modelo incondicional	179
A3.2 Segundo caso de modelo incondicional	180
A3.3 Tercer caso de modelo incondicional	181
A3.4 Cuarto caso de modelo incondicional	182
Índice de tablas	185
Índice de figuras	187



RESUMEN

Este estudio reporta el análisis multinivel efectuado a los resultados de México en la escala global de Ciencias en PISA 2006, tomando como marco teórico al Modelo de Calidad Educativa desarrollado por el INEE.

La metodología de análisis sigue esquemas utilizados en investigaciones *ex post facto*, para los que se plantearon preguntas de investigación sobre las que se formulan hipótesis relativas a cada dimensión de calidad. Las variables independientes que integran a cada dimensión de calidad, identificadas por medio de análisis factorial exploratorio y corroboradas con ecuaciones estructurales, se combinaron utilizando modelos jerárquicos de tres niveles (estudiante, escuela, región) donde la variable dependiente es el desempeño de los estudiantes en la prueba de Ciencias. Las variables independientes (o variables explicativas) fueron levantadas por medio de un cuestionario contextual que es parte del proyecto PISA 2006, incluyendo variables individuales (sexo o género, nivel socioeconómico, expectativas educativas), familiares (estructura, recursos educativos disponibles en el hogar) y escolares (nivel socioeconómico, calidad de los recursos, características del profesorado). Algunas variables de nivel regional no contenidas en la base de datos de PISA, se tenían disponibles en otros trabajos y referencias nacionales, principalmente proyectos publicados por el INEE (por ejemplo, indicadores de recursos económicos destinados a educación por estado, desarrollo humano, PIB *per cápita*, cobertura educativa, entre muchas más).

El uso de los modelos jerárquicos permitió identificar la asociación de las variables y reportar la varianza dentro de cada nivel de anidamiento para asociarla con las diferencias de los individuos o grupos que integran a cada nivel.

Se perfilaron tres objetivos principales para el estudio:

- 1) Contar con elementos que permitan explicar los patrones y organización de los datos de desempeño de los sustentantes, en función de elementos de información de tipo socioeconómica y académica, en tres niveles de anidamiento.
- 2) Producir unidades explicativas en forma de modelos de tipo cuantitativo y cualitativo, con énfasis en la interpretación de los elementos de información hacia las competencias que muestran los sustentantes.

- 3) Generar modelos que sugieran esquemas de mejora educativa en los diversos niveles de agregación, con recomendaciones y sugerencias para uso potencial de varios tipos de usuarios.

Como punto de partida se compartió el énfasis que la prueba PISA 2006 tuvo por explicar el desempeño de los estudiantes en el área de Ciencias, dejando como áreas menores Lectura y Matemáticas, pero como se sabía de las posibles limitaciones en datos nacionales que tiene la base de datos del proyecto, principalmente por su contexto internacional, se tomó el acuerdo de aprovechar otros datos procedentes de publicaciones disponibles en el INEE, atendiendo al conocimiento que se tiene del sistema educativo y la situación del país.

El estudio se mostró muy ambicioso desde un principio, no solamente había expectativas de producir resultados y sugerencias para mejorar el desempeño en la prueba de Ciencias, sino también porque se consideraba importante formular nuevas preguntas, identificar problemas emergentes y proponer o sentar las bases para otros análisis. Así, el estudio debía dar pie a nuevas preguntas y brindar sugerencias útiles para completar los cuestionarios de contexto de PISA, para hacer estudios posteriormente con las bases de datos, para aportar metodología a los proyectos del INEE y hasta sugerir o modificar alguno de los indicadores nacionales de calidad.

Por lo anterior, el proyecto se volvió un estudio de calidad educativa con los resultados de PISA 2006, más que un análisis multinivel con las bases del proyecto PISA.

Para la realización del proyecto se consideraron algunas actividades poco usuales en este tipo de estudios. Normalmente el investigador establece su muestra y los instrumentos apropiados para la investigación que está desarrollando, de conformidad con las hipótesis que se desea probar. En el presente estudio, en cambio, la información fue proporcionada por el proyecto PISA 2006, por lo que algunos aspectos de la calibración del instrumento no estaban disponibles, lo cual generó la necesidad de verificar la calidad de la prueba y de los reactivos, para garantizar el uso nacional de las medidas propuestas en la base de datos de PISA.

En la propuesta inicial se esperaba realizar un análisis multinivel con los datos de PISA. Faltaba definir el tipo de análisis y el marco teórico del estudio, con el riesgo de realizar *un estudio más* con modelos jerárquicos lineales, el primero disponible sobre la prueba PISA 2006 en México. Se percibió que esta idea no tenía un objetivo de investigación claro, por lo que había que ir mucho más allá de este propósito y se buscó una línea de investigación para evitar la tentación de construir numerosos modelos aprovechando una base de datos tan rica como la del proyecto PISA, pretendiendo *encontrar algo* a partir de las corridas de un programa de análisis multinivel. Durante el proceso para establecer las preguntas de investigación y las hipótesis, se encontró la línea de desarrollo —proporcionada por los trabajos disponibles en la literatura y por documentos del mismo INEE— mediante el marco teórico del Modelo de Calidad Educativa del INEE, con el cual se aclararon algunos aspectos no definidos o desarrollados en forma parcial, inclusive mezclados entre las dimensiones de calidad. Se obtuvo una múltiple ventaja al emplear este modelo: definir la línea de investigación, formular las preguntas e hipótesis de trabajo, proponer una base para estudios posteriores y llegar a proponer esquemas de mejora como se tenía previsto en los objetivos del proyecto.

Al esbozar el estudio, también se pensó en completar el análisis multinivel con análisis factoriales exploratorios (componentes principales) y confirmatorios (ecuaciones estructurales), por lo que fue de capital importancia la presencia del Modelo de Calidad, porque dio total sentido al estudio factorial; se revisaron las variables que participan en los componentes principales, encontrando una buena correspondencia con las dimensiones de calidad y, posteriormente, confirmando su influencia en las variables latentes propuestas por dichas dimensiones.

Los resultados son muy elocuentes de la asociación entre variables de la base de datos y el desempeño en Ciencias, y se traducen en recomendaciones a nivel del estudiante y su familia, del docente y de la escuela, de las autoridades educativas estatales y nacionales.

Presentación

Este nuevo reporte, que el Instituto Nacional para la Evaluación de la Educación pone al alcance de los lectores interesados en la educación, es una muestra más del avance en la calidad técnica del trabajo del INEE y, en particular, del que lleva a cabo la Dirección de Proyectos Internacionales y Especiales (DPIE). Esta es el área responsable en nuestro país del programa de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) conocido por las siglas PISA (*Programme for International Student Assessment*).

Es sabido que las evaluaciones PISA se aplican cada tres años, a partir del año 2000, y sus resultados llaman la atención de las autoridades, los educadores y el público en general cada vez más. Sin embargo, es frecuente que la lectura de los resultados de PISA, como la de otras evaluaciones, sea superficial, y se fije, sobre todo, en un dato de menor importancia, que es el lugar que ocupa un país o región en los ordenamientos que se pueden hacer con base en los promedios de las puntuaciones alcanzadas por los estudiantes evaluados.

Aunque llamativa, la aparente claridad de esa información es poco útil, porque el hecho de ocupar un lugar detrás de otros países en un ordenamiento no necesariamente es indicio de una situación preocupante, ya que depende de con quién se haga la comparación, y de la distancia que separe a los países comparados, lo que no suele tenerse en cuenta en los análisis usuales.

Por ello, el INEE ha insistido en que es de mayor importancia considerar las proporciones de los jóvenes evaluados que se sitúan en los diversos niveles de competencia o desempeño que identifican las pruebas, porque permite hacer juicios sobre lo adecuado o inadecuado de la situación educativa de un país en forma, independiente del lugar que ocupe en los ordenamientos.

Además de lo anterior, es de gran relevancia explorar las posibles razones por las que los estudiantes evaluados alcanzan o no niveles adecuados de competencia. La información que se recaba mediante los cuestionarios que se aplican con las pruebas permite hacer este tipo de exploración, teniendo siempre conciencia de que las con-

clusiones a las que se llegue se deberán tomar con cautela, puesto que no contarán con la base más sólida para llegar a interpretaciones en términos causales, que sólo ofrecen diseños de investigación más complejos, como los de tipo experimental.

Con la salvedad anterior, la rica información que ofrecen los estudios de PISA, junto con otros datos sobre el sistema educativo con los que se cuenta, pueden aprovecharse para hacer estudios que vayan más allá de la simple descripción de los resultados.

La participación de México en PISA ha pasado por varias etapas: en el año 2000, antes de la creación del INEE, fue muy pobre, limitándose a traducir en forma simple las pruebas recibidas y a aplicarlas a la muestra mínima exigida por la OCDE, sin intervenir en el diseño de los instrumentos, ni hacer un análisis propio de los resultados.

En el año 2003, el recién creado INEE asumió la responsabilidad de PISA en nuestro país, pero ya no pudo participar en el diseño de las pruebas, puesto que ya estaban terminadas y a punto de aplicarse. Pese a ello, la participación fue más importante, en especial porque la muestra que se utilizó fue mucho mayor, para poder dar resultados por entidad federativa y tipo de servicio, como se hizo en el informe que el INEE difundió al mismo tiempo que la OCDE liberó el informe internacional.

Aunque modesto, ese informe fue el primero que se hizo en México sobre tan importante proyecto internacional, e incluyó la información de la aplicación del año 2000, que en su momento no se había difundido. Otro elemento novedoso fue la preparación de una obra para propiciar que los maestros de bachillerato y secundaria aprovecharan las pruebas PISA para repensar su forma de trabajar en el aula, de manera que los alumnos desarrollen competencias complejas como las que evalúa PISA. Gracias al interés de la Secretaría de Educación Pública, se editaron 250 mil ejemplares de la obra, titulada *PISA para Docentes. La evaluación como oportunidad de aprendizaje*, que se hicieron llegar a todas las escuelas del país de los niveles mencionados, antes de la aplicación de PISA 2006.

En esta última, la participación de México en PISA, por conducto del INEE, fue más activa: se comenzó a intervenir en el diseño de las pruebas; se hizo una traducción de alta calidad, con innovaciones reconocidas por la OCDE; se utilizó nuevamente una muestra ampliada para tener resultados por entidad y servicio; se realizó una aplicación adicional a una muestra especial de alumnos del primer grado de enseñanza media superior, de cualquier edad; y se difundió un informe nacional muy completo, nuevamente, al mismo tiempo que el informe internacional. Para apoyar a los maestros, además, se han preparado tres volúmenes, titulados *PISA en el Aula* (sobre cada una de las competencias evaluadas: lectora, matemática y científica), que aparecerán en forma simultánea a ésta.

El informe que ahora se presenta es un avance más, cuya importancia conviene resaltar. Se trata de un esfuerzo notable de análisis de los resultados de PISA 2006, a la luz de la información derivada de los cuestionarios de contexto, y añadiendo datos sobre el sistema educativo no derivados de la aplicación de PISA, pero disponibles gracias al trabajo del INEE, en especial de la Dirección de Indicadores Educativos.

Como los lectores podrán apreciar, el trabajo es de gran complejidad técnica, ya que se utilizan técnicas estadísticas avanzadas. Además del modelo de Rasch, se hace

uso de los llamados *Modelos Lineales Jerárquicos* (HLM, por las siglas en inglés de *Hierarchical Linear Models*), así como de *Modelos de Ecuaciones Estructurales* (SEM, por las siglas de *Structural Equations Models*). El estudio, además, hace aportaciones novedosas en lo que se refiere a la forma en que se manejan los modelos multinivel.

La aportación más interesante es la inclusión de un marco teórico sobre el concepto de calidad educativa. Se utiliza el Modelo de Calidad del INEE, que da sentido a todo el trabajo de investigación, desde las preguntas y las hipótesis de trabajo, hasta el conjunto de productos que se obtienen de los análisis. Las variables utilizadas en el análisis se conciben así como indicadores de las dimensiones de eficacia, eficiencia, equidad, pertinencia, e impacto, que no son otras que las que comprenden el concepto de calidad del INEE. Esta forma de abordar el uso de los modelos multinivel difiere de otros estudios, en particular los de tipo exploratorio que carece de un marco teórico específico.

Otras innovaciones del estudio son las siguientes:

- La inclusión de variables externas a PISA. Se utilizan datos que ha recopilado el INEE y que complementan el panorama de variables explicativas para el Modelo de Calidad. Esas variables incluyen el rezago a nivel individual y el PIB a nivel regional.
- La definición de algunos puntos metodológicos cuyo uso no es claro en los trabajos de PISA, como la intervención de las ponderaciones, el proceso de imputación incluyendo una variable muda y el efecto asociado dentro de la ecuación de regresión, o la justificación del Índice de Estatus Socioeconómico y Cultural (ESCS por sus siglas en inglés). Se hizo un estudio específico para la validación de la prueba, incluyendo calibración de ítems y verificación de las escalas; este análisis proporciona la seguridad de que se trabaja con datos apropiados para estudiar el desempeño de los estudiantes de México.
- La identificación de las variables por cada nivel y su agrupación, por medio de un análisis factorial, confirmado con el *Modelo de Ecuaciones Estructurales*. Este enfoque no se suele presentar; a pesar de ser parte de la metodología que permite validar la pertinencia de los cuestionarios de contexto.
- El planteamiento del análisis multinivel sobre las variables explicativas relativas a cada dimensión latente de la calidad, de acuerdo con las agrupaciones factoriales, en lugar de utilizar todo el conjunto de variables observadas o disponibles en la base de datos de PISA 2006.

Al ir más allá de la descripción de los resultados, y explorar posibles explicaciones de ellos, el informe será de interés especial para las autoridades educativas, que encontrarán en él pistas sugerentes para el diseño de políticas de mejora educativa.

Por otra parte, el informe pone en evidencia el avance de la competencia técnica del personal de la DPIE del INEE, con lo que se cumple el segundo propósito fundamental por el que es importante que México participe en proyectos de evaluación como PISA: aprender de las mejores prácticas internacionales, para desarrollar la capacidad técnica de nuestro personal. El primer propósito, complementar el conocimiento de la calidad de la educación mexicana al compararla con la de los sistemas educativos de los países más avanzados del mundo, se cumple también más ampliamente con este informe.

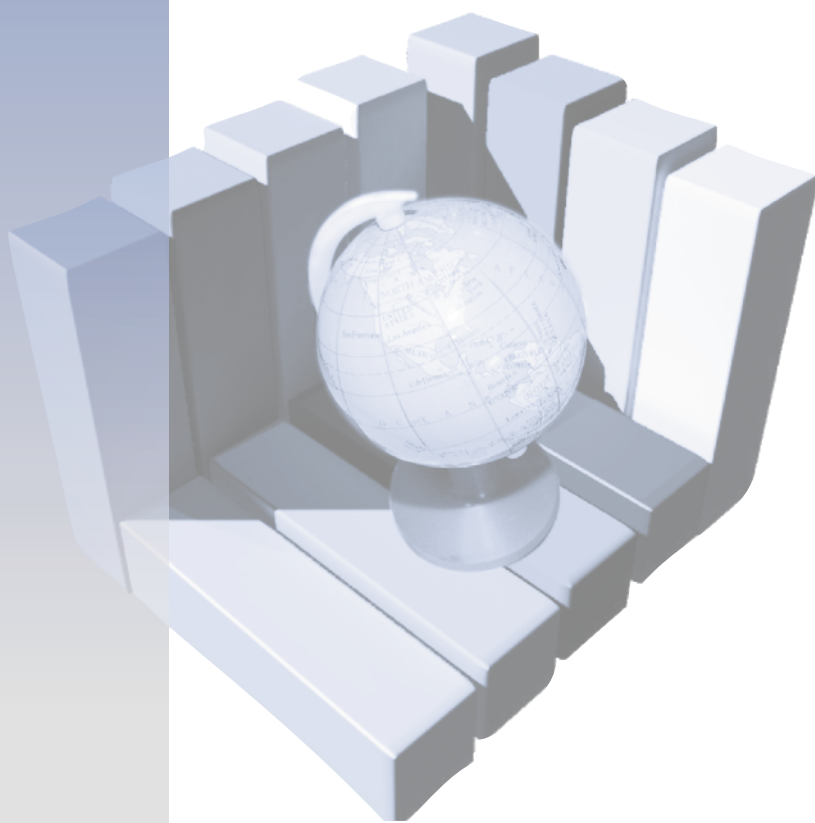
Agradezco al personal de la DPIE, a María Antonieta Díaz, directora de la DPIE por haber impulsado el estudio y, en especial, a los miembros de la subdirección de Análisis y Aprovechamiento de Resultados Internacionales quienes, bajo la conducción de Agustín Tristán, colaboraron en la preparación de este informe, así como al propio Doctor Tristán quién fungió como coordinador del estudio, y a Ismael Flores por su asesoría para el tratamiento de la muestra y la base de datos. No se debe olvidar al resto del personal, en especial de la Dirección de Relaciones Nacionales y Logística, así como al de las Áreas Estatales de Evaluación, que hicieron posible la aplicación de PISA.

Felipe Martínez Rizo

Noviembre de 2008

Capítulo 1

El Modelo de Calidad Educativa del INEE



Capítulo I. El Modelo de Calidad Educativa del INEE

I.1 Necesidad de un Modelo de Calidad de la educación

Es una tarea difícil tratar de identificar una definición de calidad educativa que tenga amplia aceptación en el contexto educativo y sociopolítico de un país. Las propuestas que se han publicado a lo largo de los años, reflejan el interés de los factores que participan en los procesos educativos y la dificultad de integrar todas las facetas que involucra la mejora de la educación.

Como un trabajo de reflexión que parte de experiencias muy variadas, de aportaciones de los consejeros y de un minucioso análisis de referencias nacionales e internacionales, el INEE ha postulado el Modelo de Calidad Educativa (INEE, 2004 y 2006; Martínez, R.F., 2004 y 2005; Martínez, R.F. y col., 2007). Se trata de un modelo multidimensional que considera los componentes del sistema educativo dentro de un enfoque sistémico de tipo CIPP (Contexto-Insumo-Proceso-Producto), pero que trasciende las cuatro facetas sistémicas y se concreta en ejes o dimensiones de calidad.

A partir del modelo CIPP (Stufflebeam y Anthony, 1987), pueden generarse esquemas de metaevaluación con el propósito de asegurar la calidad de los servicios educativos, definiendo metas para perfeccionar el trabajo evaluativo (Scriven, M. 1969; Stufflebeam, D.L. y Shinkfield J. A., 1981).

El enfoque sistémico con el modelo CIPP, considera cuatro componentes principales:

Contexto: Integra el conjunto de variables que reflejan la realidad social que permite o impide el desarrollo educativo, por medio de funciones y condiciones de operación. Dentro de las variables de contexto se contemplan la forma en que los recursos disponibles (materiales, económicos, humanos) deben atender a las necesidades educativas de la sociedad, debidamente caracterizada por sus diferentes estamentos, niveles socioeconómicos y capital cultural, su distribución a lo largo y ancho del país.

Insumos: Identifica todos los recursos que recibe el sistema educativo y que favorecen su funcionalidad. Como parte de los insumos se contemplan variables tales como: materiales, financieras, tecnológicas, humanas o de cualquier otra índole.

Procesos: Se refiere a las operaciones necesarias para la organización escolar, el liderazgo y las relaciones humanas involucradas en el funcionamiento de los planteles.

La planeación, gestión y evaluación son algunos de los procesos que inciden en la organización y que deben considerar diversos procesos de tipo psicopedagógico, de investigación y de trabajo en el aula. En un nivel diferente, los procesos se extrapolan dentro de cada estado y el país en su conjunto. En el concepto del INEE, la regulación de los procesos se refiere a la definición y manejo de los objetivos o propósitos del sistema educativo; para orientar y organizar los insumos en la búsqueda de los productos deseables.

Productos: Distingue o describe los resultados del sistema educativo: desde los logros de los estudiantes hasta el impacto que se tiene en la sociedad; desde el ámbito de influencia de la escuela hasta el bienestar que se produce en una zona, un estado o todo el país, y que puede incidir en una mejora de la calidad de vida de la población en general.

El uso de un modelo CIPP se ha presentado en una buena recopilación de referencias dada por Murillo et al. (2007), o el trabajo de De la Orden y col. (1997). En su mayoría, los trabajos enfatizan *la eficacia*, considerada una dimensión más operacional que las otras, que permite aproximarse a la calidad o medirla en función de algunos indicadores objetivos. Los estudios de los autores presentan esquemas y modelos publicados previamente en la literatura o nuevas combinaciones producto de consideraciones metodológicas actuales, por lo que pueden servir como punto de partida para establecer nuevos modelos bajo el enfoque sistémico.

Utilizar el enfoque de metaevaluación o el modelo CIPP en los procesos de control de calidad, tiene como problema de partida la definición misma de *calidad*. Las definiciones son muy diversas, debido a que ésta es un concepto multifactorial, que no puede generalizarse al tener que incluir atributos específicos de un entorno o contener un sesgo de diseño, dependiendo de la forma en que se consideren la eficiencia, la eficacia, el impacto, la equidad, la importancia, la pertinencia, la vigencia o el crecimiento sostenible, la utilidad, la factibilidad, la ecología, los derechos humanos y la visibilidad, entre otros criterios. De todas las formas en que se defina la calidad se espera tener un resultado objetivo que retroalimente al proceso de evaluación.

A pesar del notable esfuerzo de recopilación y de organización, el uso de los modelos CIPP no resuelve el problema de definición de calidad, porque todas las consideraciones hechas y los conceptos que integran a la calidad no están definidas de manera multidimensional. En los trabajos se encuentra que el concepto de eficacia, eje principal de los estudios, se mezcla o confunde con la eficiencia y con la calidad misma. Esto hace que la clasificación no sea correcta, produciendo deficiencias operativas propias de la indefinición y de la intersección de variables no excluyentes entre sí.

Fuera de lo atractivo que puede ser el modelo CIPP, resulta insuficiente para definir el concepto de calidad, porque los mismos componentes de dicho modelo CIPP pueden producir resultados diferentes, dependiendo de los propósitos o dimensiones deseables definidos por las políticas educativas de un país. La gran variedad de esquemas o modelos CIPP que se han propuesto para la calidad educativa dan testimonio de esta falta de precisión.

La necesidad de establecer debidamente el concepto de calidad es un paso que ya realizó el INEE al señalar ejes transversales del modelo sistémico. Se trata de un conjunto de atributos multidimensionales que permite definir criterios de evaluación mediante elementos de mayor concreción, por lo tanto más específicos y apropiados al quehacer educativo.

1.2 Modelo de Calidad Educativa del INEE

Para el INEE (Martínez R.F., 2005, 2007), la calidad del sistema educativo es “la cualidad que resulta de la integración de las dimensiones de pertinencia y relevancia, eficacia interna, eficacia externa, impacto, suficiencia, eficiencia y equidad”. Puede observarse que en esta definición se consideran algunas dimensiones y subdimensiones que podrían ubicarse como contexto, proceso, insumo y, finalmente, como producto, de ahí el interés de esta definición de calidad, porque proporciona un marco teórico más preciso que el modelo sistémico para los fines educativos nacionales.

“Estas dimensiones de la calidad se derivan de una relación de congruencia entre los componentes del sistema educativo entre sí: la pertinencia, la relevancia, la eficacia externa y el impacto se derivan de la relación de los objetivos y productos del sistema con las necesidades del entorno; la eficacia interna se desprende de la relación entre productos y objetivos; la eficiencia depende de la relación entre insumos o recursos y productos o resultados. La equidad se deriva de la adecuada distribución de recursos y productos educativos.”

El concepto de calidad educativa del INEE se define como relativo y dinámico. Se indica que es relativo porque hay elementos propios de cada país, región o contexto particular; que dimensionan las expectativas que se hacen sobre los logros educativos. Se dice que es dinámico porque representa más un proceso que un momento, de tal modo que la calidad es un objetivo a perseguir aunque no sea factible de alcanzar; ya que siempre pueden considerarse otros elementos que propicien la mejora de la calidad. Se dice por ello que la calidad no es un estado sino una tendencia asociada con el concepto de superación y mejora continua.

La definición del Modelo de Calidad Educativa del INEE puede englobarse en las cinco dimensiones que se describen a continuación en forma operacional o concreta:

Eficacia (interna y externa): Se refiere a la medida en que se logran los propósitos educativos y, generalmente, se traduce en indicadores de rendimiento o de éxito. A este respecto se considera de importancia el conseguir la más alta proporción posible de estudiantes que alcanzan los objetivos (eficacia externa) por grupos de edad o por tipo de escuela, altas tasas de titulación o de eficacia terminal, así como bajas tasas de reprobación o abandono escolar (eficacia interna). También se puede referir a la obtención de los niveles esperados de desempeño o dominio de los objetivos de aprendizaje, tanto para el uso académico como para su posible incidencia en otros niveles escolares o en la vida cotidiana y el trabajo.

Eficiencia y suficiencia: Hace alusión a la forma de utilizar los recursos disponibles (humanos, materiales, de uso del tiempo y económicos), que deben estar suministrados en cantidad suficiente para los propósitos que se persiguen. Cuando el empleo de los recursos es apropiado para darles el mejor aprovechamiento posible, se dice que se alcanza la eficiencia.

Equidad: Se refiere al logro de los propósitos por el mayor número posible de personas, tomando en cuenta las desigualdades de origen (individual, familiar, escolar, estatal, regional) para adaptar la metodología o procesos a cada situación ofreciendo las adaptaciones o apoyos requeridos para los propósitos citados. La interpretación de esta dimensión se refleja en esquemas de intervención psicopedagógica, cultural o socioeconómica. En el ámbito escolar se pueden considerar acciones como una mayor oferta educativa, disponibilidad de horarios y docentes, acceso a fuentes de consulta y mayores oportunidades educativas en general. A este respecto, Gago y Mercado (1995) indican que no debe confundirse la equidad con el paternalismo, tampoco recomiendan la reducción de los niveles o estándares de calidad o a la práctica de establecer criterios diferenciados entre grupos, lo cual representa de suyo una práctica discriminatoria.

Pertinencia y relevancia: Señalan la forma en que los resultados corresponden con los postulados, expectativas y necesidades sociales (incluyendo los rubros cultural, político, económico, social) y del ámbito académico (uso de los conocimientos, las habilidades y otras competencias desarrolladas a través de las disciplinas). Sobre esta base se puede llegar a que el currículo sea adecuado a las necesidades individuales de los alumnos (pertinencia), pero sin dejar de lado a las necesidades de la sociedad (relevancia), tanto para conducir a competencias que apoyen el bienestar social y la productividad económica, sin olvidar las facetas de desarrollo integral y humanista, tales como el respeto de los derechos humanos, la apreciación de los valores democráticos, el interés por el desarrollo de la ciencia, la protección al medio ambiente, la conservación y enriquecimiento de la diversidad cultural.

Impacto: Se manifiesta cuando los aprendizajes que desempeñan los alumnos tienen una vigencia temporal de largo plazo, con miras amplias para transformar su apreciación de la vida y de la sociedad, sustentados en los valores de libertad, solidaridad, tolerancia y respeto a las personas. Si se toma en cuenta que la manifestación de los aprendizajes debe producir el bienestar social, entonces la persona deberá estar en posibilidad de emplear sus competencias en la vida cotidiana, convirtiéndolo en un ciudadano completo, de tal manera que la formación trasciende hacia las competencias para la vida, permaneciendo vigentes a lo largo del tiempo.

Se puede observar que estas dimensiones siguen un concepto clasificatorio que garantiza que se trata de un conjunto completo o exhaustivo de categorías excluyentes, de modo que ninguna de estas dimensiones es suficiente por sí sola para caracterizar de manera integral la calidad educativa.

Igualmente se observa que se trata de un Modelo de Calidad multidimensional que no enfatiza a una sola de las dimensiones, por ello su representación es un pentágono equilátero (similar al modelo propuesto por Gago y Mercado, 1995), a diferencia de otras visiones que enfatizan una dimensión o un propósito particu-

lares, como refieren De la Orden y col. (1997), con esquemas de calidad que la consideran como un fenómeno excepcional, una perfección o coherencia, el ajuste a un propósito.

Figura 1.1 Modelo de Calidad Educativa del INEE



Un diseño no equilátero, desfavorecería una de las dimensiones, produciendo paradigmas de calidad con interpretaciones no esperadas o no deseables, como podrían ser modelos que persiguen la eficacia sin cuidar la equidad, modelos que se antojan muy equitativos pero que no usan los recursos de manera eficiente, o modelos muy eficientes con un bajo impacto o trascendencia.

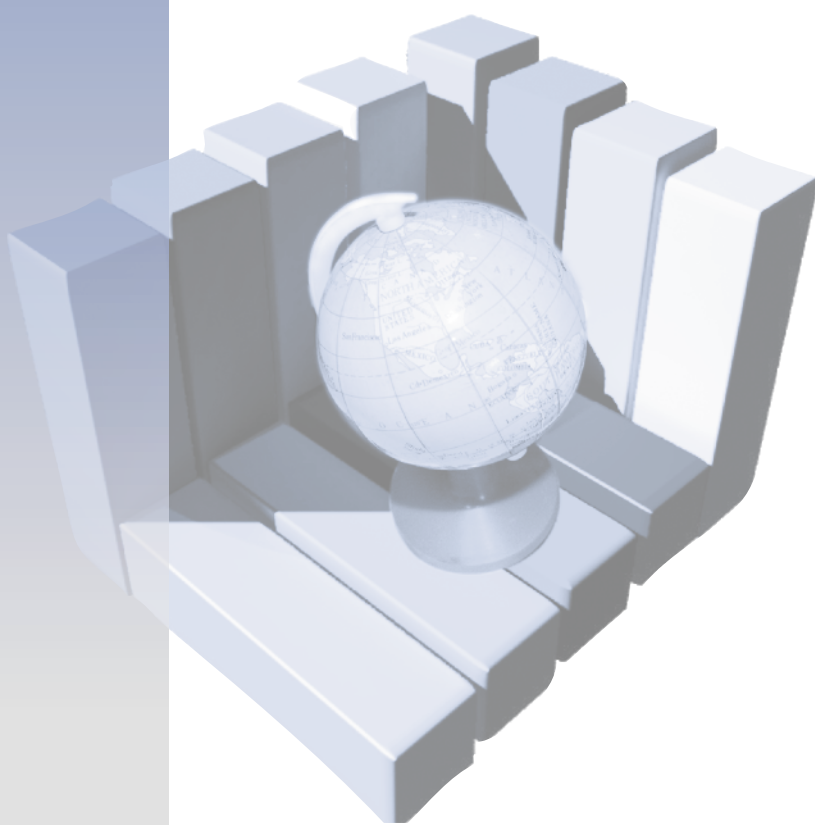
Adicionalmente a las cinco dimensiones de la calidad, se ha contemplado una variable que representa la situación en la cual se ubica el estudiante, resultado de aspectos internos y externos que pueden explicar condiciones que determinan ciertos resultados en el desempeño de la prueba PISA, como se explica en el capítulo 3.

El Modelo de Calidad Educativa del INEE ofrece un marco teórico y una base de información suficiente (Anexo 1) para la realización del presente estudio. Las variables recabadas por los cuestionarios de contexto del proyecto PISA, junto con otros datos procedentes de fuentes nacionales, se organizan en tres niveles: (1) por estudiante (factores de la familia y factores propios del estudiante); (2) por escuela, pudiendo distinguirse aspectos de aula (factores del docente y factores de clima y organizacionales del salón) y aspectos del plantel (sobre el personal directivo, condiciones psicopedagógicas y organización de la escuela misma) y (3) región (principalmente aspectos socioeconómicos de las regiones, cuestiones educativas como la cobertura de estudiantes y la infraestructura).

El Modelo de Calidad Educativa del INEE sirve como marco teórico para el análisis de los datos procedentes de los cuestionarios de PISA. El estudio se realiza con modelos multinivel para estimar el efecto de las variables en el desempeño en Ciencias que tienen los estudiantes que participan en la prueba PISA.

Capítulo 2

Metodología del estudio



Capítulo 2. Metodología del estudio

2.1 Descripción general del proyecto PISA

2.1.1 Presentación

El Programa para la Evaluación Internacional de los Estudiantes (*Programme for International Student Assessment, PISA*), promovido y organizado por la OCDE (Organización para la Cooperación y el Desarrollo Económicos) es un estudio comparativo y periódico en el que pueden participar los países miembros o no miembros (asociados) de la OCDE.

La información derivada de PISA es un indicador del nivel de aptitud o habilidad de los estudiantes en su propio país, en comparación con otros países participantes; ayuda a identificar las fortalezas y debilidades de los sistemas educativos nacionales; permite encontrar patrones del desempeño de los estudiantes entre países; y sobre todo, detecta qué factores se asocian al éxito educativo más allá de establecer comparaciones entre resultados de forma aislada.

El propósito principal de PISA es evaluar en qué medida los estudiantes de 15 años han adquirido conocimientos y habilidades esenciales para participar plenamente en la sociedad, y hasta qué punto son capaces de extrapolar lo aprendido para aplicarlo a situaciones novedosas, tanto del ámbito escolar como extraescolar. La evaluación mira hacia adelante, se interesa en la capacidad de los estudiantes para aplicar sus conocimientos y habilidades al enfrentar los retos de la vida real, más que en medir hasta qué punto dominan un plan de estudio o currículo escolar. Esta orientación va más allá de medir la capacidad de los estudiantes para reproducir lo aprendido, en su lugar la prueba PISA explora las habilidades de los jóvenes para continuar aprendiendo durante toda su vida, su capacidad de aplicar lo aprendido y tomar decisiones en diversos tipos de ambientes (escolares y no escolares).

2.1.2 Los ciclos

Otra característica importante de PISA es la regularidad de su aplicación. PISA se ha establecido como un programa trianual y en cada ejercicio contiene tres áreas de

evaluación, pero hace énfasis en un área prioritaria que ocupa dos terceras partes de las preguntas de la evaluación. En la aplicación del año 2000, el énfasis fue en el área de Lectura; en el 2003, en Matemáticas; y en el 2006, en Ciencias; ésta última sirvió de enfoque para el análisis objeto del presente estudio.

2.1.3 Población objetivo

Con el fin de garantizar que los resultados sean comparables entre países, PISA evalúa poblaciones semejantes. Debido a que existen diferencias entre países en cuanto a la naturaleza y alcance de la educación preescolar; la edad de entrada a la enseñanza obligatoria y la estructura de los sistemas educativos; los grados escolares no son comparables internacionalmente. En consecuencia, para comparar válidamente los resultados educativos de unos países y otros, se opta por definir la población objetivo con referencia a una edad determinada.

De esta forma, la evaluación de PISA comprende a estudiantes de entre 15 años tres meses y 16 años dos meses de edad al momento de la evaluación, que estén inscritos en una institución educativa a partir del séptimo grado; se excluyen a los estudiantes que se encuentren en sexto grado o por debajo. La población objetivo de México en PISA 2006, se conformó con los estudiantes nacidos en 1990 que al momento de la evaluación se encontraban inscritos en secundaria o en la educación media superior.

2.1.4 Instrumentos

Se aplican dos tipos de instrumentos escritos: *los cuadernillos de conocimiento y los cuestionarios de contexto*.

Los *cuadernillos de conocimiento* están diseñados conforme a un esquema matricial para asegurar una mayor cobertura de contenidos, sin necesidad de que todos los estudiantes respondan la totalidad de reactivos. Bajo esta estructura, los cuadernillos se integran por diferentes módulos de las áreas de evaluación (Ciencias, Matemáticas y Lectura) que incluyen diversas unidades de reactivos.

Los *cuestionarios de contexto* usados en PISA están dirigidos a los estudiantes y al director de la escuela, y se utilizan para el análisis de los resultados, ya que proporcionan información de las características de los estudiantes y la escuela.

En particular, el cuestionario del estudiante contiene información sobre: antecedentes económicos, sociales y culturales de los estudiantes y sus familias; actitudes ante el aprendizaje, hábitos escolares y estilos de vida en el ámbito familiar; aspectos del proceso de enseñanza-aprendizaje.

El *cuestionario del director* contiene información sobre características de la escuela, entre ellas los recursos humanos y materiales, el financiamiento y las prácticas administrativas, aspectos sobre el contexto académico y el nivel de compromiso de las familias en la educación de sus hijos.

2.1.5 Muestra

Se utilizan muestras representativas que oscilan entre 4 mil 500 y 10 mil estudiantes de un mínimo de 150 escuelas de cada país, de manera que es posible realizar inferencias para el país en su conjunto, no así para regiones o estados. El proyecto permite que los países participantes soliciten una sobre muestra, con el propósito de disponer de una mayor representatividad respecto a cierto estrato de la población. Siguiendo esta opción, México solicitó, para la aplicación de 2006, una sobre muestra para producir resultados significativos a nivel entidad, por lo que se evaluó a 30 mil 971 estudiantes en mil 140 escuelas.

En las modalidades recabadas por el proyecto PISA 2006, que engloba a la secundaria y al nivel medio superior, solo se excluyeron de la base de datos los estudiantes inscritos en Secundarias para Trabajadores (104 personas) y en Escuelas de Capacitación para el Trabajo (517 personas), debido a que la representatividad dentro de la muestra resultaba dudosa y a que existieron duplicaciones de sujetos.

2.1.6 Países participantes

En PISA 2006, la participación ascendió a 57 países (30 miembros de la OCDE y 27 asociados). Se evaluó a aproximadamente a 400 mil estudiantes seleccionados al azar y quienes representan a cerca de 20 millones de jóvenes de 15 años de las escuelas de esas naciones.

2.2 Descripción de la metodología para el estudio

El estudio que se desarrolla aquí utiliza datos procedentes de la prueba PISA para verificar e interpretar el Modelo de Calidad Educativa del INEE. La base de datos del proyecto está integrada por índices que fueron calculados con base en las respuestas de los estudiantes y las autoridades de la escuela a las preguntas contenidas en los cuestionarios de contexto. El interés está puesto en las puntuaciones que obtienen los alumnos en la prueba de Ciencias, bajo la presentación de valores plausibles y se incluyen datos socioeconómicos, demográficos y académicos a nivel región procedentes de estudios previos realizados por el INEE.

La metodología seguida en este estudio de investigación *ex post facto*, identifica escenarios para cada dimensión de calidad a partir preguntas de investigación y la formulación de hipótesis sobre el comportamiento de las variables que forman a la dimensión, obtenidas a través de análisis factoriales exploratorios y confirmadas con ayuda de ecuaciones estructurales. En cada escenario, la dimensión de calidad se concibe como una variable mediadora latente, asociada con aquéllas del componente factorial (que son las variables independientes del modelo), con los valores plausibles del desempeño en Ciencias de la prueba PISA (que son las variables dependientes). Este conjunto de variables se expresa con un modelo jerárquico multinivel, con el cual se determina (1) la participación de las variables independientes sobre el desempeño de los estudiantes en la prueba y (2) la forma en que se

reparte la varianza total de las puntuaciones en Ciencias entre los niveles de anidamiento (estudiante-escuela-región).

A continuación se detalla la metodología utilizada en el modelo.

2.2.1 Escenarios para el desarrollo de las hipótesis

En este estudio se ha considerado la pertinencia de establecer un juego de hipótesis sobre las cuales se desarrollen diferentes modelos relativos a la calidad educativa. A diferencia de otros trabajos de investigación *ex post facto*, se ha considerado la posibilidad de enunciar hipótesis sobre las variables de la base de datos de PISA 2006. Este estudio no responde a las características de una investigación experimental, sino más propiamente a un estudio descriptivo, porque se cuenta con estas condiciones que acotan el enfoque del proyecto: (1) muestra definida de antemano con el fin de hacer la medición de las competencias en Ciencias usando pruebas de bloques incompletos; (2) definición de las variables contextuales por el proyecto PISA y no por hipótesis definidas por el proyecto de investigación planteado aquí; (3) inexistencia de grupos de contraste o de control.

Hay varias posibilidades para abordar un estudio como éste. Para el desarrollo de las hipótesis y los modelos de un estudio, se identifican tres posibles escenarios o enfoques, como sugiere Byrne, B.M. (2001):

1. Estrictamente confirmatorio (SC). El investigador postula un modelo con base en cierta teoría, recoge los datos de una muestra apropiada al caso y prueba el ajuste del modelo hipotético a los datos de la muestra. En este caso, los resultados pueden confirmar o descartar la hipótesis o el modelo.
2. Modelos alternativos (AM). El investigador propone varios modelos alternativos con base en la teoría. Recoge los datos apropiados al estudio y acepta el modelo que produce la mejor representación de los datos de la muestra.
3. Generación de modelo (MG). En este caso se procede de manera exploratoria a modificar y replantear un primer modelo que sirve de base, aunque no sea suficientemente satisfactorio. La exploración consiste en identificar los desajustes y llegar a un modelo que represente los datos de la mejor manera posible; es decir, se debe llegar a un modelo suficientemente significativo o explicativo y de buen comportamiento estadístico.

Byrne comenta que es muy común el tercer enfoque, especialmente porque se trata de evitar hacer una investigación para acabar rechazando el modelo que se planteó desde un principio. Esta tendencia es reportada por numerosos investigadores y meta-analistas (ver por ejemplo, Rosenthal, R., 1979; Wayne, A. y Youngs, P., 2003; Rodríguez, M.C., 2005; Else-Quest, N.M., Shibley, H.J., Hill Goldsmith, H. y Van Hulle, C.A., 2006; Seidel, T. y Shavelson, R.J., 2007; Timperley, H. y Alton-Lee, A., 2008) quienes hacen énfasis en que hay un sesgo de publicación – también conocido como problema del cajón de archivo o *file drawer problem* – por el cual es más probable que se publiquen los estudios con hallazgos estadísticamente significativos que los que no tienen estos hallazgos. Debido a este tipo de evidencia, los investigadores evitan enviar a publicación trabajos importantes pero que no muestran resultados estadísticamente significativos, perdiéndose información valiosa en el camino.

El peligro del tercer escenario es que tiene el atractivo de buscar el modelo de mejor ajuste entre las variables; es común tratar de interpretar dicho modelo ajustado como si fuera correlacional o hasta causal, sin ningún sustento. Esto es una invitación a hacer el cruce de todas las variables contra todas, hasta encontrar alguna *información* interesante para ser explicada. Por tales razones, se decidió no utilizar este tercer escenario para el presente estudio.

Puede anotarse, con base en lo anteriormente expuesto, que es común establecer modelos descriptivos de ajuste, que tratan de representar lo mejor posible el comportamiento de los datos. En este caso se corre el riesgo de construir modelos con demasiadas variables y perder parsimonia en aras del ajuste, así como establecer modelos que resultan muy complejos y que no pueden ser comprendidos fácilmente, explicados razonablemente y aprovechados en políticas nacionales, institucionales o en acciones personales.

A diferencia de los modelos exploratorios, que empiezan determinando lo que se conoce como modelo nulo (utilizado generalmente en modelos multinivel), para después enriquecerlo o complementarlo con una o varias variables explicativas, en este trabajo se presentan las hipótesis propuestas ante una pregunta de investigación, el modelo que la representa, su interpretación y conclusiones respecto de su pertinencia. Cada hipótesis puede dar lugar a uno o varios modelos, en función de un enriquecimiento por combinaciones de las variables explicativas, pero sin perder de vista el enfoque inicial.

2.2.2 Formulación de hipótesis

La formulación de hipótesis parte de la idea de que se trata de un enunciado probable sobre las variables o sus relaciones, no verificado pero susceptible de ser probado o refutado (Larroyo, F. y Cevallos, M.A., 1965). De Landsheere, G. (1976) indica que sin una hipótesis que sirva de directriz la investigación puede degenerar en una acumulación estéril de datos e información. Por su parte Kerlinger, F.N. y Lee, H. (2002), al igual que Myers, A. y Hansen, C.H. (2002) aclaran que además de que la hipótesis es establecida como enunciado que propone una relación variable, sirve de guía a la investigación que se debe realizar sobre ellas. Los autores hacen ver que un análisis como el que se realiza en este estudio, como diseño correlacional y *ex post facto*, siendo no experimental, sí puede incluir alguna hipótesis. A este respecto puede verse también el enfoque para el propósito de la investigación que sugieren Hernández, Fernández y Baptista (2004).

Existen varios criterios para la formulación de hipótesis y su presentación en un informe de investigación, pero no se cuenta con un modelo único, sucede que en los documentos de investigación pueden aparecer la hipótesis general, las hipótesis específicas, las hipótesis estadísticas o no aparecer en absoluto, dejando al lector el trabajo de reconocerla a partir de los resultados. De acuerdo con las normas APA (ver también Thompson, 1966), en este reporte de investigación se incluyen hipótesis que responden a una pregunta donde se establecen relaciones entre el problema, el diseño experimental y la propia hipótesis.

2.2.3 Metodología para el estudio

Para la realización del estudio, se siguió una metodología general que, como todo proceso de investigación, requirió de varias aproximaciones conforme se complementaban o depuraban los datos, o durante el proceso de aclarar ideas que permitieran culminar los análisis. Se deben distinguir varios cuerpos de información.

En primer lugar se tienen todos los elementos del proyecto PISA: los valores proporcionados en la base de datos o los que tuvieron que ser asignados, como es el caso de los valores omitidos que requieren de una imputación, los pesos regionales o los datos regionales para el nivel 3, así como el manejo de los valores plausibles relativos al desempeño en Ciencias de los estudiantes.

En segundo lugar están los modelos sugeridos para el análisis de los datos: análisis factorial exploratorio, modelos multivariados de ecuaciones estructurales y modelos multinivel y que requieren de *software* especializado para realizar los análisis.

Con base en dichos cuerpos de información se concreta la metodología para la producción y análisis de las hipótesis que se establecen en relación con el Modelo de Calidad Educativa del INEE. Esta metodología se ve reflejada en la presentación del análisis para cada eje o dimensión contenida en el capítulo 3 de este estudio, siguiendo este esquema:

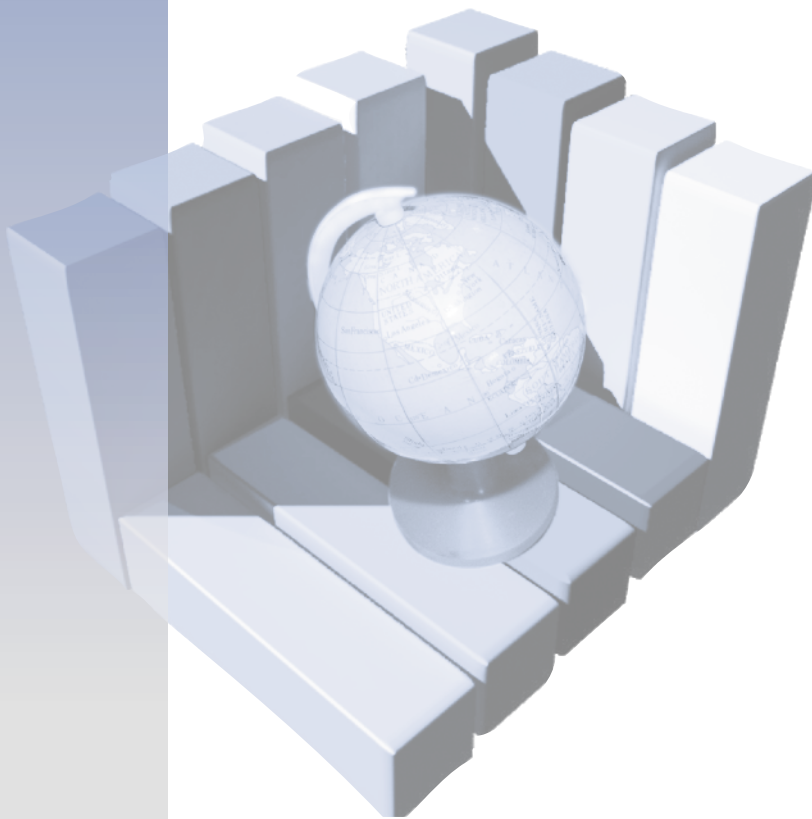
1. Preguntas de investigación. Se trata de una o varias preguntas que fijan la participación de variables del componente factorial en asociación con el desempeño de los estudiantes en Ciencias.
2. Hipótesis. Presenta los enunciados concretos para atender la pregunta de investigación. Debe recordarse que las hipótesis se van a probar con la base de datos disponible.
3. Variables explicativas. Dispone de una tabla con las variables elegidas por nivel para el eje o dimensión de calidad que se estudia. Para un mayor detalle sobre las variables deberán consultarse las tablas contenidas en anexo.
4. Resultados del análisis factorial confirmatorio. El conjunto de variables elegidas, que provienen del análisis factorial exploratorio, se combinan en un modelo de ecuaciones estructurales que relacionan variables observadas y variables latentes con la puntuación en Ciencias de PISA. Se incluye la presentación en forma de diagrama obtenido con el programa de análisis de ecuaciones estructurales, donde se indican los coeficientes estandarizados para cada relación entre variables y sus correlaciones en su caso. Se indica el índice de ajuste AGFI (índice de bondad de ajuste por sus siglas en inglés AGFI, el cual debe ser cercano a 1) para tener un elemento de decisión sobre la convergencia del modelo. El mayor interés de este análisis se ubica en el aspecto cualitativo de los coeficientes estandarizados, principalmente su magnitud y su signo.
5. Ecuaciones del modelo multinivel. A partir de la expresión del modelo de tres niveles (Estudiante-escuela-región) se llega a los resultados dispuestos en forma tabular. Dentro de las tablas se incluyen los términos fijos (intercepto y pendientes o coeficientes de las variables) y el resumen de la varianza en los términos

aleatorios; en ambos casos se incluye la significancia de los valores comparados contra un valor nulo y la participación que tiene la varianza de cada nivel en la varianza total. También se incluyen tablas que presentan el desajuste del modelo (deviance) y su comparación contra el modelo incondicional o modelo nulo, para decidir si hubo mejora en el modelo con la dimensión de calidad estudiada. Con ayuda de la comparación contra el modelo incondicional, también es posible determinar el porcentaje de la varianza que se explica en el nuevo modelo en estudio.

6. Presentación e interpretación de resultados. Los datos obtenidos se presentan una vez adaptados al criterio de la d de Cohen (Cohen, J., 1988), con lo que se puede dictaminar si se tienen diferencias significativas en los resultados de Ciencias al considerar valores extremos de las variables. Se incluyen gráficas con los resultados.
7. Conclusiones e implicaciones de los resultados. Para terminar el análisis de cada dimensión de calidad, se emiten algunas conclusiones y sugerencias para utilizar los resultados en actividades de mejora continua, mediaciones didácticas u organizacionales, en los niveles individual, escolar o regional, que podrían incidir en mejores resultados de los estudiantes.

Capítulo 3

Análisis



Capítulo 3. Análisis

3.1 Preliminares

Antes de presentar los resultados de los modelos contruidos para las dimensiones de calidad, conviene hacer la recapitulación de los análisis preliminares realizados, las agrupaciones de variables respecto al Modelo de Calidad del INEE y el análisis incondicional (o análisis nulo) que sirve de referencia para todos los demás estudios.

3.1.1 Análisis realizados

El estudio partió de la conveniencia de garantizar que los datos de PISA fueran aplicables para las comparaciones entre regiones de México, atendiendo a que las bases se plantearon inicialmente con fines de contraste internacional. En caso de encontrarse un sesgo en los datos para las regiones de México, se haría la recalibración y reconstrucción de la escala sin sesgo, además de construir el conjunto de valores plausibles para cada registro de estudiante.

También se vio la necesidad de identificar las variables para representar al Modelo de Calidad Educativa del INEE, con base en análisis multivariados donde se utilizaron los archivos de datos procedentes del proyecto PISA en los niveles de anidamiento 1 y 2 (estudiante y escuela). Para los datos del nivel de anidamiento 3 (regiones), se emplearon datos socioeconómicos, demográficos, académicos y culturales disponibles de otros estudios del INEE.

Se hizo un análisis factorial exploratorio, con rotación varimax, para identificar los posibles factores dentro de cada nivel de anidamiento y su interpretación como dimensiones del Modelo de Calidad. Este procedimiento facilitó y respaldó la elección de variables que, de otra forma, se hubiera tenido que justificar con las agrupaciones empleadas en la literatura, pero que no se asocian con el modelo del INEE. Las agrupaciones fueron discutidas en sus aspectos cualitativos entre los integrantes del grupo responsable de este estudio. Los resultados obtenidos de estas agrupaciones son interesantes, pero no deben tomarse como definitivos, porque debe recordarse que los cuestionarios contextuales y la base de datos regionales no

responden directamente a un diseño experimental para probar o refutar el Modelo de Calidad Educativa del INEE. Los datos son cualitativamente interesantes en esta etapa del estudio y pueden servir de referencia para futuros trabajos sobre esta misma base, una vez adaptado el diseño de los cuestionarios y la toma de datos en las muestras nacional y estatal.

El conjunto de variables de cada dimensión se tradujo en modelos de ecuaciones estructurales para su comprobación funcional. Los modelos que se consideran aquí fueron convergentes en el número de iteraciones y la precisión esperada, pero debe recordarse que no implican un óptimo, sino una posible solución al problema de identificar relaciones entre variables. El uso de las ecuaciones estructurales se justifica desde el punto de vista cualitativo, por ser confirmatorio de los componentes factoriales, además de que proporciona coeficientes que pueden interpretarse en cuanto a magnitud y sentido para medir los efectos que tienen las variables independientes sobre las variables dependientes (observadas o latentes).

El procedimiento seguido hasta aquí tiene la ventaja de que los modelos estructurales sirven como guía para los análisis multinivel, especialmente al configurar el conjunto de variables más trascendentes que pueden emplearse como explicativas en los modelos jerárquicos. Como podrá verse a continuación, en el nivel 1 no se tiene clara la participación de la eficacia, que sí se manifiesta de forma nítida en el nivel 3. La eficiencia, la equidad y la pertinencia participan en los niveles 1 y 2. Finalmente, todo apunta a que las dimensiones de impacto se ubican en el nivel 1.

El desempeño de los estudiantes participantes en Ciencias en el proyecto PISA se representó con un modelo multinivel o modelo jerárquico lineal en una estructura en tres niveles, donde los estudiantes (Nivel 1) están agrupados o anidados en escuelas (Nivel 2) y éstas a su vez en regiones (Nivel 3). Los análisis se realizan sobre cada una de las dimensiones del Modelo de Calidad Educativa del INEE por separado.

Es importante resaltar que las observaciones sobre las variables no cognitivas (actitudes, intereses y valores) no son concluyentes, debido a la naturaleza de las variables y a la forma en que se construyeron con base en las respuestas de los estudiantes a preguntas de los cuestionarios de contexto de PISA. Dichas observaciones solamente deben tomarse como conjeturas; como propuestas de posibles líneas de investigación o de exploración para trabajos futuros, para los cuales se deberá recabar nueva información, previa definición operativa de las variables de interés y con una mayor precisión en la forma de estimar los índices.

3.1.2 Análisis de validez de la escala de la prueba PISA 2006

Uno de los puntos de mayor importancia de toda prueba es su validez a demostrarse por los responsables del diseño utilizando muy diversas fuentes de evidencias. Un análisis detallado de las evidencias de validez de la prueba PISA queda fuera del propósito de este trabajo, sin embargo se puede anotar que los estándares internacionales exigen que el proceso que se sigue para el diseño y las validaciones de las pruebas incluyan el trabajo de numerosos especialistas, quienes garanticen la validez facial o de

aspecto de la prueba, la validez de contexto y de contenido de los ítems, la validez de constructo y de criterio por medio de jueceo y en forma experimental y, por último, la validez de los cuestionarios contextuales.

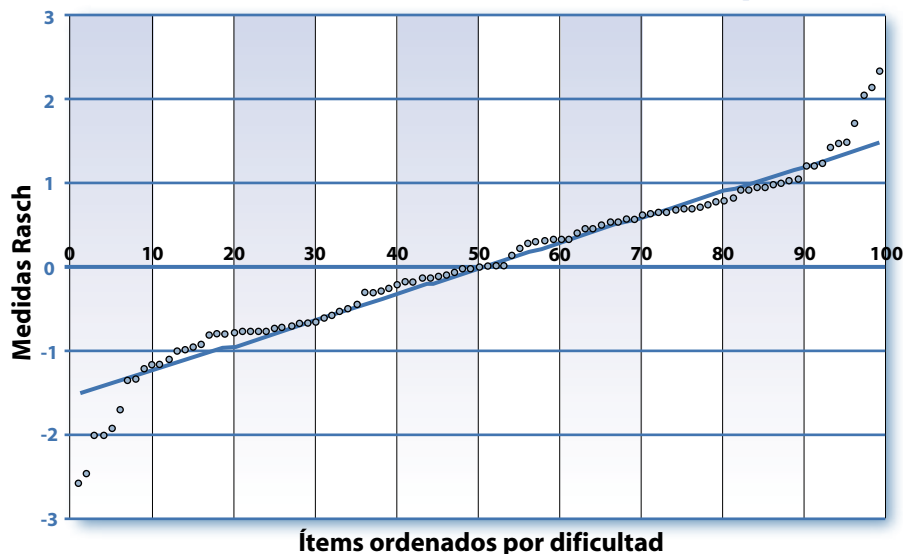
Una parte importante de las evidencias de validez que no se había explorado y de la cual no se tenían datos era en lo referente a la calibración de los ítems y al análisis técnico de la prueba. Para este proyecto se planteó como primer paso garantizar que la escala de la prueba estuviera correctamente construida, con lo cual se evitaría presentar resultados correctos para las comparaciones internacionales, pero sesgados para su uso nacional. Se calibraron los ítems de la prueba con el modelo de Rasch, para obtener los parámetros de medida y ajuste, así como la evidencia de la validez de escala que produce el instrumento. Se tenía previsto que en caso de que este proceso de recalibración indicara que la prueba no era apropiada para las comparaciones en México, se procedería a preparar una actualización de la base de datos con la escala corregida, incluyendo el conjunto de cinco valores plausibles para la puntuación en Ciencias de cada estudiante.

El proceso de recalibración resultó muy interesante, porque se encontró que hay muy pocos reactivos deficientes por bajas correlaciones o por desajuste. Sólo se encontraron cinco ítems con deficiencias de poca importancia, en particular dos con correlaciones punto-biserial menores a 0.2 pero superiores a 0.1, otros dos tuvieron ligero desajuste al modelo de Rasch y sólo uno tuvo ambas deficiencias simultáneamente. La presencia de estos cinco ítems con deficiencias menores no es de importancia significativa como para representar un problema en la calidad de la prueba.

Una vez que se demostró que los reactivos son aceptables en general, se procedió a verificar la validez de escala, observándose que los reactivos de la prueba siguen muy de cerca la distribución establecida por la recta de diseño teórica que se traza de -1.5 a +1.5 lógitos¹ (Tristán y Vidal, 2007) que garantiza la distribución uniforme de las dificultades de los ítems, denominada TDL[-1.5,+1.5,0] (figura 3.1). El desajuste de la escala que produce la prueba PISA en Ciencias es inferior a 2% respecto de la recta de diseño.

¹ El lógito es una unidad de medida de los modelos logísticos y, en particular, del modelo de Rasch. Esta medida se obtiene con el logaritmo natural del cociente aciertos/fallas del desempeño de los sujetos. El cociente citado se denomina momio.

Figura 3.1 Distribución de la dificultad de los ítems de Ciencias en la prueba PISA



Por la cantidad tan limitada de reactivos deficientes y por la validez de la escala obtenida, se tomó el acuerdo de no recalibrar las medidas de los sujetos, con lo cual no fue necesario estimar nuevos valores plausibles para la base de datos de PISA 2006.

La media de las medidas de los sujetos se ubica en -0.57 lógitos (esto indica que el desempeño tiene una deficiencia mediana respecto del instrumento, que está centrado en 0 lógitos²), con una desviación estándar de 1.03 lógitos, muy cercana al valor esperado de 1.0.

La confiabilidad de la prueba completa es de 0.77, con una separación de 1.81³, lo cual es aceptable para un instrumento de más de 100 ítems.

El mapa de Wright o mapa de ítems-personas se representa en la figura 3.2. A la izquierda se tiene la distribución de frecuencias de las medidas de los sujetos, a la derecha se tiene la de los ítems, ambas se ubican en una escala que corre de -5 a +4 lógitos. Los valores positivos de la parte superior del mapa indican un mayor dominio en las personas y un ítem más difícil. Las letras que aparecen cerca del eje de la escala, indican la ubicación de la media (M), la primera desviación estándar (S) y la segunda desviación estándar (T). En este mapa de Wright se puede apreciar la distribución de tipo gaussiana de los puntajes de las personas en el área de Ciencias, ante una prueba que tiene una distribución muy uniforme, principalmente en el intervalo de validez de escala de -1.5 a +1.5 lógitos.

² De acuerdo con la escala de medidas que se obtiene con el modelo de Rasch, una media positiva indica mayor dominio de parte de los sujetos, en tanto que una media negativa indica una deficiencia en los sujetos. El valor absoluto de la diferencia de la media con respecto del valor centrado en 0 puede catalogarse de esta manera: nula=diferencia menor a 0.25 (1/4 de lógito), pequeña= diferencia menor a 0.5 (1/2 lógito), mediana= diferencia menor a 1 lógito, grande=diferencia superior a 1 lógito.

³ La separación indica el número de estratos discernibles que pueden encontrarse en la escala a partir del conjunto de ítems de una prueba; a mayor valor de separación se tiene una mayor confiabilidad en el sentido clásico de consistencia interna de la prueba.

Figura 3.2 Mapa de Wright de la prueba de Ciencias de PISA

TABLE 1.1 Todo PISA (precalificado)		cogni3.sal Jun 9 21:30 2008
INPUT: 30971 PERSONS, 103 ITEMS MEASURED:		30932 PERSONS, 102 ITEMS, 2 CATS 3.58.1
<hr/>		
MAP OF PERSONS AND ITEMS		
MEASURE	PERSONS +-- ITEMS	MEASURE
<more>		<rare>
4	. + XXX	4
	.	
	.	
	.	
	.	
	.	
3	. +	3
	.	
	.	
	.	
	. X	
	. # X	
2	. +T X	2
	. #	
	. ### X	
	. ## T XXX	
	. ##### X	
	. ### XX	
1	. ##### +S XXXXXXXXX	1
	. ##### XXXX	
	. ##### XXXXXXXXX	
	. ##### S XXXXXXXX	
	. ##### XXXXXXXX	
	. ##### XX	
0	. ##### +M XXXXXXXX	0
	. ##### XXXXXXXX	
	. ##### XXXX	
	. ##### M XXXX	
	. ##### XXXXXXXX	
	. ##### XXXXXXXX	
-1	. ##### +S XXXXXXXX	-1
	. ##### XXXX	
	. ##### XX	
	. #####	
	. ##### S X	
	. #####	
-2	. ##### +T XXXX	-2
	. #####	
	. #####	
	. ## X	
	. # T X	
	. #	
-3	. ### +	-3
	.	
	.	
	.	
	.	
-4	. +	-4
	.	
	.	
	.	
	.	
-5	. ## +	-5
<less>	PERSONS +-- ITEMS	<frequent>
EACH '#' IN THE PERSON COLUMN IS 82 PERSONS: EACH '.' IS 1 TO 81		

En conclusión, el análisis de la validez de la escala de la prueba de Ciencias de PISA proporciona la garantía de que los valores plausibles no solamente sirven para las comparaciones internacionales, sino que también definen una escala apropiada para las comparaciones en México, con lo cual no se necesita hacer un ajuste por sesgo a los datos disponibles.

3.1.3 Agrupación de variables

Los resultados del análisis factorial conducen a definir agrupaciones de las variables de la base de datos⁴. Se observará que los modelos parten de las hipótesis propuestas

⁴ En algunos casos es muy difícil argumentar si la variable pertenece al rasgo latente, por ejemplo la dimensión de eficacia, o si se trata de una variable de contexto para la variable dependiente. Esta argumentación queda fuera del propósito de este trabajo y es parte de las dificultades que han encontrado otros investigadores que han estudiado la eficacia y la calidad educativas.

en relación con el desempeño en términos de los ejes o dimensiones de la calidad, de tal forma que no hay mezclas entre variables, para poder explicar el efecto que se tiene de cada grupo de variables explicativas en las puntuaciones de los estudiantes. La descripción de las variables se tiene en anexo y las agrupaciones en las dimensiones de calidad se tienen en las tablas 3.1 a 3.3.

Las cinco dimensiones de la calidad pueden representarse como variables latentes en los tres niveles de anidamiento. Esto se aprecia claramente en el nivel 1, donde cada variable aparece predominantemente en un componente factorial. En el nivel 2 algunas variables pueden interpretarse en dos componentes como IRATCOMP y CL-SIZE. En el nivel 3 todas las variables se estandarizaron con media cero y desviación estándar 1; el primer componente es preponderante (que se asimiló a la dimensión de eficacia) y deja pocas opciones de interpretación para los otros factores.

Tabla 3.1 Resultados del análisis factorial exploratorio del nivel 1 (valores absolutos)

VALORES MÁXIMOS ABSOLUTOS									
Variables	0.949	0.802	0.644	0.844	0.767	0.878	0.713	0.856	0.919
PARED	0.949	0.068	0.059	0.013	0.001	0.070	0.040	0.009	0.002
HISCED	0.943	0.073	0.064	0.013	0.002	0.062	0.039	0.002	0.004
PESCS	0.903	0.382	0.042	0.004	0.026	0.045	0.059	0.047	0.019
ESCS	0.894	0.411	0.047	0.006	0.028	0.046	0.064	0.047	0.019
ESCS2	0.768	0.222	0.009	0.002	0.011	0.029	0.009	0.147	0.024
HISEI	0.737	0.222	0.025	0.012	0.013	0.031	0.105	0.017	0.015
HOMEPOS	0.510	0.802	0.028	0.011	0.064	0.008	0.019	0.13	0.037
WEALTH	0.484	0.729	0.078	0.005	0.061	0.023	0.001	0.136	0.101
HEDRES	0.392	0.652	0.067	0.025	0.098	0.075	0.05	0.090	0.014
SC02Q01	0.107	0.555	0.025	0.004	0.019	0.108	0.043	0.165	0.025
CULTPOSS	0.298	0.535	0.260	0.021	0.014	0.082	0.051	0.069	0.124
ENVAWARE	0.094	0.142	0.644	0.039	0.063	0.077	0.382	0.001	0.057
SCIEEFF	0.099	0.139	0.629	0.060	0.100	0.007	0.224	0.020	0.072
JOYSCIE	0.069	0.113	0.611	0.036	0.399	0.007	0.014	0.049	0.083
SCSCIE	0.050	0.022	0.599	0.207	0.213	0.001	0.005	0.015	0.010
SCIEFUT	0.068	0.173	0.592	0.077	0.275	0.028	0.210	0.024	0.051
CARINFO	0.082	0.112	0.591	0.223	0.091	0.024	0.220	0.022	0.130
SCHANDS	0.036	0.022	0.123	0.844	0.066	0.016	0.021	0.027	0.027
SCAPPLY	0.019	0.034	0.181	0.842	0.124	0.002	0.046	0.034	0.024
SCINVEST	0.039	0.028	0.117	0.827	0.071	0.034	0.128	0.015	0.010
GENSCIE	0.063	0.112	0.134	0.037	0.767	0.031	0.108	0.001	0.071
PERSCIE	0.007	0.006	0.325	0.064	0.763	0.012	0.039	0.016	0.005
RESPDEV	0.055	0.054	0.126	0.081	0.518	0.063	0.428	0.026	0.019
CARPREP	0.024	0.003	0.235	0.346	0.515	0.019	0.094	0.012	0.007
NIVEL	0.084	0.079	0.01	0.028	0.031	0.878	0.157	0.039	0.048
REZAGO	0.069	0.080	0.004	0.017	0.024	0.847	0.084	0.027	0.036
ENVOPT	0.051	0.007	0.044	0.048	0.002	0.055	0.713	0.004	0.015
ENVPERC	0.025	0.059	0.032	0.030	0.316	0.057	0.474	0.084	0.362
URBANORURAL	0.123	0.098	0.016	0.021	0.028	0.054	0.066	0.856	0.003
GENERO	0.033	0.051	0.033	0.026	0.074	0.076	0.049	0.005	0.919
	EQUIDAD	EFICIENCIA	IMPACTO1	PERTINENCIA	IMPACTO2	SITUA	IMPACTO3	SITUA	SITUA

Tabla 3.2 Resultados del análisis factorial exploratorio del nivel 2 (valores absolutos)

Valores máximos absolutos				
Variables	0.78	0.81	0.86	0.64
XESCS	0.78	0.17	0.15	0.20
SC02Q01	0.76	0.00	0.06	0.26
SCMATEDU	0.68	0.00	0.14	0.03
IRATCOMP	0.47	0.14	0.00	0.50
NIVEL	0.14	0.81	0.27	0.01
PROPCERT	0.01	0.77	0.02	0.02
CLSIZE	0.03	0.58	0.17	0.45
ENVLEARN	0.14	0.06	0.86	0.05
SCIPROM	0.14	0.21	0.83	0.08
STRATIO	0.16	0.05	0.08	0.64
URBANORURAL	0.45	0.12	0.02	0.63
IRRESPCURR	0.34	0.35	0.15	0.24
	Eficiencia o equidad	eficacia	relevancia	eficiencia

Tabla 3.3 Resultados del análisis factorial exploratorio del nivel 3 (valores absolutos)

Valores máximos absolutos						
VARIABLES	0.982	0.853	0.696	0.743	0.606	0.404
ZMATPRIV	0.982	0.027	0.055	0.091	0.039	0.151
ZESPUBL	0.971	0.053	0.105	0.113	0.105	0.142
ZMATPUBL	0.963	0.170	0.104	0.106	0.070	0.129
ZESPPIV	0.954	0.107	0.009	0.199	0.180	0.074
ZCOMPUPUBL	0.944	0.164	0.238	0.008	0.144	0.068
ZCOBERTURA	0.939	0.245	0.230	0.057	0.048	0.007
ZGRESOLARIDAD	0.908	0.129	0.385	0.005	0	0.108
ZPCONTMUJER	0.889	0.074	0.421	0.146	0.022	0.073
ZCOMPUPRIV	0.888	0.364	0.100	0.167	0.112	0.169
ZMEDIAPVREGION	0.888	0.298	0.073	0.033	0.074	0.332
ZPIPERCAPITA	0.867	0.357	0.273	0.042	0.207	0.046
ZMARGINALIDAD	0.863	0.203	0.442	0.026	0.061	0.122
ZIDH	0.862	0.148	0.442	0.054	0.192	0.013
ZESPGRAL	0.861	0.105	0.202	0.333	0.310	0.025
ZMEDIAPVESTADO	0.861	0.289	0.131	0.164	0.221	0.286
ZEDADNORMA	0.370	0.853	0.241	0.149	0.155	0.175
ZDOCLICENC	0.324	0.840	0.269	0.099	0.194	0.264
ZIMIGRANTE	0.031	0.821	0.217	0.160	0.488	0.116
ZAPROBTELE	0.186	0.744	0.086	0.594	0.184	0.131
ZPTERMTELE	0.595	0.702	0.205	0.175	0.282	0.002
ZGASTOXALUMNO	0.378	0.634	0.199	0.123	0.606	0.182
ZPTERMGRAL	0.498	0.515	0.476	0.166	0.481	0.029
ZDOCEXPER	0.422	0.131	0.696	0.224	0.409	0.321
ZDOCCARRERA	0.466	0.474	0.633	0.080	0.387	0.019
ZANALFAB	0.755	0.232	0.604	0.040	0.081	0.050
ZGRXANALF	0.759	0.225	0.601	0.038	0.103	0.021
ZMATRICULA	0.440	0.021	0.412	0.743	0.263	0.122
ZBIBLIOT	0.538	0.505	0.276	0.581	0.192	0.065
ZAPROBTEC	0.760	0.204	0.034	0.487	0.182	0.331
ZAPROBGRAL	0.821	0.174	0.040	0.460	0.207	0.199
ZDOCCAPAC	0.844	0.443	0.114	0.013	0.061	0.273
ZPCONTTOTAL	0.826	0.122	0.533	0.135	0.025	0.008
ZGRSOREANALF	0.803	0.152	0.485	0.287	0.111	0.046
ZPCONTHOMBRE	0.769	0.240	0.571	0.127	0.019	0.088
ZGASTOFAEB	0.730	0.396	0.447	0.226	0.197	0.144
ZMATGRAL	0.715	0.365	0.480	0.251	0.249	0.027
ZPTERMTEC	0.689	0.144	0.533	0.209	0.120	0.404
	Eficacia					

3.2 Modelo incondicional

El primer modelo de tres niveles, totalmente incondicional⁵, es decir, sin variables predictoras en los niveles, sirve para identificar la forma en que las medidas del desempeño en Ciencias se distribuyen entre los tres niveles citados: estudiante, escuela y región⁶. Este modelo se utiliza como una referencia para el resto del análisis pero, a diferencia de otros estudios, no se pretende ir enriqueciéndolo por la inclusión de más variables puesto que cada modelo está definido de antemano por la selección de las variables explicativas reportadas en el análisis factorial y confirmadas con el análisis de ecuaciones estructurales.

El modelo incondicional describe el desempeño en Ciencias de un estudiante cualquiera sin utilizar variables explicativas. Es una función donde aparece la media de los resultados de cada escuela (intercepto), más un error aleatorio:

$$Y_{ijk} = \pi_{0jk} + e_{ijk}$$

Donde:

Y_{ijk} es el desempeño en Ciencias de un estudiante i en la escuela j de la región k
 π_{0jk} es la media de puntuaciones en Ciencias en la escuela j de la región k
 e_{ijk} es el efecto aleatorio que se tiene en el estudiante.

En este modelo, la puntuación real de cada estudiante se obtiene al sumar el efecto aleatorio a la puntuación media de su escuela. Este efecto aleatorio es la desviación o diferencia que tiene cada persona respecto de la media del grupo del cual forma parte. Las desviaciones se distribuyen en forma normal, con media cero y desviación estándar σ (o varianza σ^2).

Para el nivel 2 la media de la escuela se puede representar por una ecuación similar:

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$

en esta ecuación se tiene que:

β_{00k} es la media de puntuaciones en la región k
 r_{0jk} es el efecto aleatorio o desviación que tiene cada escuela respecto de la media regional.

⁵ También se conoce como modelo nulo o modelo vacío. Este modelo reporta a la variable dependiente sin variables que establezcan una asociación con ella. Como resultado de este modelo se reporta la media global de desempeño en Ciencias como valor base (intercepto) y las tres fluctuaciones aleatorias dependientes de cada nivel.

⁶ En el anexo se presenta la manera en que fueron definidas las regiones para el estudio.

De nuevo, los efectos aleatorios del segundo nivel se consideran con una media nula y una varianza τ_k . La puntuación real de cada escuela se obtiene al sumar un efecto aleatorio a la puntuación media de todas las escuelas en la región.

Finalmente, para el tercer nivel, la variabilidad entre regiones se representa con la ecuación:

$$\beta_{00k} = \gamma_{000} + u_{00k}$$

donde

γ_{000} es la media general de puntuaciones de toda la muestra de sustentantes

u_{00k} es el efecto aleatorio o desviación que tiene cada región respecto de la media general.

Estos efectos aleatorios se distribuyen en forma normal con media 0 y varianza τ_β y se interpretan de manera similar a los anteriores. Ahora la puntuación de cada región se obtiene al sumar su efecto aleatorio a la media de todas las regiones.

Al construir el modelo multinivel se puede escribir el conjunto de ecuaciones en una sola (también denominado modelo mixto), quedando de esta forma:

$$Y_{ijk} = \gamma_{000} + u_{00k} + r_{0jk} + e_{ijk}$$

donde:

Y_{ijk} es el desempeño en Ciencias de un estudiante i en la escuela j de la región k

γ_{000} es la media general (efecto fijo) de puntuaciones de toda la muestra de sustentantes:

u_{00k} es el efecto aleatorio que tiene cada región

r_{0jk} es el efecto aleatorio que tiene cada escuela.

e_{ijk} es el efecto aleatorio que se tiene en el estudiante.

Esta presentación es la elegida para este estudio porque tiene la ventaja de que es más breve, además de que incluye los elementos fijos y aleatorios del modelo sin pasar por coeficientes intermedios de cada nivel.

Cada elemento o unidad de análisis se interpreta en función del nivel, pudiendo ser un estudiante, una escuela o una región; de este modo se tienen estos datos de partida:

- Total de unidades en el nivel 1: 30 mil 350 estudiantes
- Total de unidades en el nivel 2: un mil 088 escuelas
- Total de unidades en el nivel 3: 7 regiones

De acuerdo con la metodología de análisis propuesta por el proyecto PISA, se deben emplear los cinco valores plausibles de la puntuación de Ciencias y la ponderación o peso regional de los datos de la muestra en la población analizada. En el anexo 3 se muestran los resultados de realizar análisis sin utilizar valores plausibles

ni ponderaciones, con ello se justifica el uso de estas ponderaciones hasta llegar al modelo que se propone aquí y cuyos resultados se presentan en la tabla 3.4.

Tabla 3.4 Resultados del modelo incondicional

Efecto fijo		Coeficiente	Error estándar	t	Grados de libertad	P-valor
Puntuación media global en Ciencias	γ_{000}	418.17	10.73	38.97	6	0.000

Efecto aleatorio			Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
Nivel	Unidad							
1	Estudiante	e_{ijk}	58.46	3416.98				49.40
2	Escuela	r_{ojk}	55.35	3063.17	1081	24936.46	0	44.30
3	Región	u_{00k}	21.00	441.09	6	198.68	0	6.40

Desajuste (Promedio)	336665.13	← Modelo de referencia
----------------------	-----------	------------------------

Puede observarse que al utilizar los cinco valores plausibles y la ponderación por región, las participaciones de los efectos aleatorios de los niveles 1 y 2 en la varianza total están muy balanceados (fracción de varianza al nivel de estudiantes de 49.4% y 44.3% a nivel de escuelas), mientras que las regiones solo tienen 6.4% de participación en la varianza.

En este caso, la media global es de 418.17 puntos de la escala PISA. La desviación estándar de los efectos aleatorios son: 58.46 para el nivel de los estudiantes, 55.35 para el nivel de las escuelas y para las regiones es 21.00 puntos de la escala PISA.

El valor de desajuste (deviance) de 336665 se empleará como referencia para dictaminar la mejora que proporcionan los modelos de los siguientes análisis. Este desajuste es el promedio de los modelos realizados con los cinco valores plausibles incluyendo los ponderadores regionales.

3.3. Modelo de la situación del estudiante

3.3.1 Pregunta de investigación

Para este estudio se define como la situación del estudiante a un conjunto de variables propias del alumno o de su entorno que inciden en los resultados del desempeño de la prueba PISA. La situación del estudiante se define con las variables cuya participación resulta de interés para el estudio pero que no se agrupan en un componente factorial de forma clara. A partir de ella se puede establecer esta pregunta de investigación:

¿Existen variables de la situación del estudiante que producen diferencias estadísticamente significativas en el desempeño en Ciencias de los escolares evaluados en PISA 2006?

3.3.2 Hipótesis

Las variables del nivel I incluidas en los componentes factoriales (consideradas como variables que representan la situación del estudiante) son de importancia para el estudio, porque producen diferencias estadísticamente significativas en las puntuaciones de Ciencias de la prueba PISA.

3.3.3 Variables explicativas

Hay un conjunto de variables recabadas en el cuestionario de contexto de PISA que no se englobaron claramente en un componente factorial, sin embargo se consideran de importancia para el modelado y se agruparon como variables de la situación del estudiante. Resulta evidente que otras variables de cualquiera de los niveles podrían ser consideradas como situacionales para el estudiante, la escuela o la región, pero el criterio que se adoptó es que si una variable quedaba incluida dentro de uno de los componentes del análisis factorial, entonces se tomaría en cuenta al modelar la dimensión con la cual se asoció cualitativamente.

Tabla 3.5 Variables explicativas del modelo de la situación del estudiante

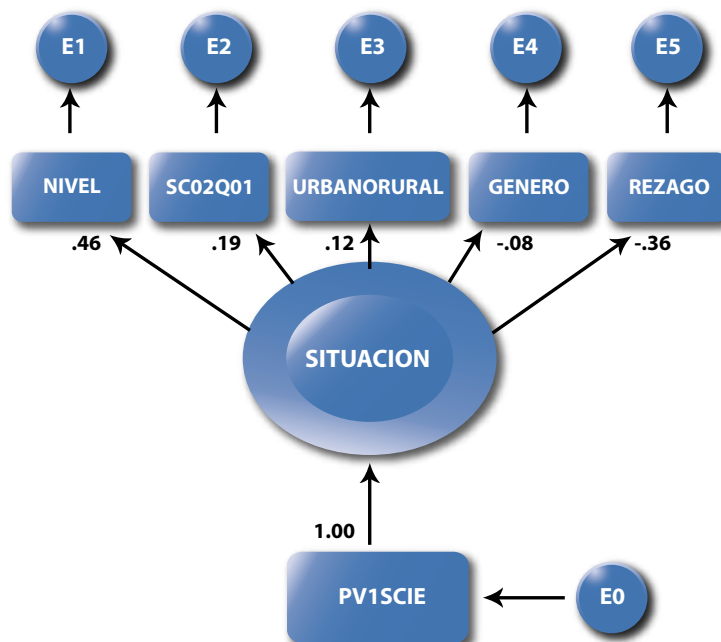
Nivel	Variables	Descripción
Variables de nivel I	NIVEL	Nivel de estudios del estudiante: 0=Secundaria, 1=Educación media superior
	SC02Q01	Tipo de sostenimiento de la escuela: 0=Público, 1=Privado
	URBANORURAL	Tipo de comunidad donde se encuentra la escuela: 0=Rural, 1=Urbano
	GENERO	Género del estudiante. Codificación: 0=Hombre, 1=Mujer.
	REZAGO	Se considera que un estudiante es rezagado si cumplió años en el periodo de enero-agosto de 1990 y está inscrito en Secundaria. Codificación: 0=Sin rezago, 1=Rezagado

3.3.4 Resultados del análisis factorial confirmatorio

Las variables se modelaron asociándolas a una latente mediadora denominada SITUACION (figura 3.3), que presenta condiciones de referencia en las cuales se encuentra el estudiante, por ejemplo: la región en la que vive, el nivel educativo en el cual está inscrito, si estudia en una escuela Pública o Privada, en ambiente Urbano o Rural y el efecto que se tiene del género. También se incluye el rezago, que puede incidir en proporcionar una situación favorable o desfavorable al desarrollo del estudiante; se trata de una condición que puede resultar crucial para que un estudiante tenga

mejores resultados tanto académicos como en su vida misma. El ajuste es bueno con AGFI=0.957 (Ver Anexo A2.4).

Figura 3.3 Modelo estructural de la situación del estudiante



AGFI=0.957

En este modelo solamente intervienen variables de nivel 1⁷. Los coeficientes que relacionan a cada variable independiente con la variable latente mediadora SITUACION, indican un mayor efecto positivo en las variables NIVEL (Secundaria y Educación media superior) y SC02Q01 que corresponde con el tipo de escuela (Pública o Privada), así como REZAGO con una participación negativa, de tal modo que un mayor rezago deberá producir un menor desempeño en Ciencias. De menor interés se tiene URBANORURAL que reporta la ubicación del centro escolar. Es interesante apreciar que es negativo el coeficiente que relaciona a la situación con GENERO, lo cual es desfavorable hacia las mujeres y confirma la tendencia hallada en otros estudios.

3.3.5 Ecuación del modelo multinivel

La ecuación del modelo jerárquico lineal que se obtiene con esta combinación de variables es:

$$PVSCIE_{ijk} = \gamma_{000} + \gamma_{100} NIVEL_{jk} + \gamma_{200} SC02Q01_{ijk} + \gamma_{300} URBANORURAL_{ijk} + \gamma_{400} GENERO_{ijk} + \gamma_{500} REZAGO_{ijk} + r_{0jk} + u_{00k} + e_{ijk}$$

⁷ El modelo también se simuló considerando el componente de varianza en la variable latente, con un menor ajuste que el del modelo mostrado.

No se incluye explícitamente la variable REGION, que queda contenida en el tercer nivel y cuya contribución se aprecia en el coeficiente aleatorio u_{00k} .

El modelo considera los valores plausibles y la ponderación ajustada a nivel regional, como se señala en el anexo 2.

Los coeficientes γ_{ijk} corresponden con el intercepto (que sirve como valor base para la puntuación de Ciencias de los estudiantes) y con las pendientes del modelo para cada una de las variables explicativas del primer nivel (que sirven como indicador de la fluctuación de la variable dependiente conforme hay un cambio unitario en la variable independiente).

Todas las variables explicativas empleadas en este modelo intervienen de manera estadísticamente significativa, como se presenta en la tabla 3.6.

Tabla 3.6 Modelo multinivel de las variables que integran la situación del estudiante

Efecto fijo		Coeficiente	Error estándar	t	Grados de libertad	P-valor
Interceptos						
Puntuación base global en Ciencias	γ_{000}	371.57	8.54	43.51	6	0.000
Pendientes						
NIVEL	γ_{100}	44.97	3.71	12.12	243	0.000
SC02Q01	γ_{200}	29.32	6.09	4.82	2018	0.000
URBANORURAL	γ_{300}	32.96	5.28	6.25	65	0.000
GENERO	γ_{400}	-15.86	1.72	-9.23	4375	0.000
REZAGO	γ_{500}	-14.95	2.94	-5.09	54	0.000

Efecto aleatorio			Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
Nivel	Unidad							
1	Estudiante	e_{ijk}	57.76	3336.42				61.00
2	Escuela	r_{0jk}	44.33	1965.28	1081	13841.11	0.0	35.93
3	Región	u_{00k}	12.95	167.64	6	89.71	0.0	3.07

No.	Nivel		Componente de varianza en el modelo nulo	Componente de varianza en este modelo	Diferencia de varianza %
1	Estudiante	e_{ijk}	3416.98	3336.42	2.40
2	Escuela	r_{0jk}	3063.17	1965.28	35.80
3	Región	u_{00k}	441.09	167.64	62.00

Desajuste (Promedio) 335523.77 Mejora respecto del modelo de referencia 0.34 %				
Diferencia respecto del modelo nulo	1141.36	Diferencia en grados de libertad	5	Diferencia significativa

Este modelo mejora en forma significativa el desajuste del modelo incondicional, pero no refleja cambios en el primer nivel (las variables de SITUACION explican⁸ 2.4% de la varianza del primer nivel), pero inciden de forma más notable en los otros dos niveles. En el segundo nivel (escuela) se produce 35.8% de explicación en la varianza, pero todavía queda un total de 35.93% de varianza que habrá que analizar con otras variables a nivel escuela. Sin haber hecho intervenir más que variables del nivel I, esta combinación explica 62% de la varianza del nivel 3, que ocupa tan solo 3.07% de la varianza total de las puntuaciones de los estudiantes en Ciencias.

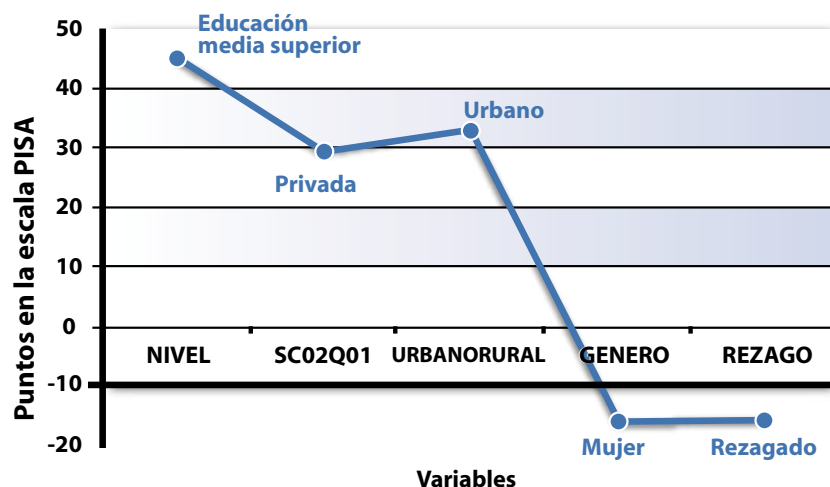
3.3.6 Presentación e interpretación de resultados

A diferencia del modelo incondicional de la sección 3.2, el intercepto γ_{000} que representa el valor base del modelo, no corresponde con la media global debido a dos causas: en primer lugar a que en este modelo se incluyeron variables explicativas y en segundo lugar, porque no se solicitó al programa el centrado de los valores respecto de la media global. Esta elección se hizo para poder utilizar la escala de cada variable sin tener que referirla a una media de cada una de ellas.

Tomando en cuenta los valores codificados en la base de datos se representan las diferencias en la figura 3.4. Los estudiantes de Educación media superior tienen en promedio un desempeño superior a los de Secundaria en 44.97 puntos de la escala PISA (equivalente a 59.4% de la desviación estándar de las puntuaciones de PISA). Los estudiantes de las escuelas privadas obtienen en promedio 29.32 puntos por arriba de los estudiantes de las escuelas públicas, un valor muy similar a la diferencia que se tiene entre escuelas urbanas y rurales de 32.96 (38.7 a 43.5% de la desviación estándar respectivamente). Las mujeres están desfavorecidas casi en 16 puntos respecto de los hombres, un valor semejante al que se aprecia entre estudiantes rezagados y regulares casi con 15 puntos (cada una de estas variables representan aproximadamente 20% de la desviación estándar de las puntuaciones en Ciencias de la prueba PISA).

⁸ La varianza del nuevo modelo es menor que la del modelo incondicional en el nivel de interés. La diferencia entre esta varianza y la que se tenía con el modelo incondicional es producto de la intervención de las variables en el modelo, por ello se conocen como *variables explicativas* y la diferencia de varianzas (o su proporción porcentual) es lo que se denomina como *explicación* respecto de la varianza del modelo incondicional. No debe confundirse este término con la palabra *explicar* en sentido coloquial.

Figura 3.4 Diferencias en desempeño de Ciencias por variables de la situación del estudiante



El interés del análisis se centra en las diferencias entre valores o categorías extremas, para lo cual se emplea el coeficiente d^* que se obtiene del criterio propuesto por la d de Cohen (1998), que establece que una diferencia es considerada como pequeña (en este caso estadísticamente significativa, ver anexo A2.8) si entre dos puntuaciones se tiene un valor menor a un *valor crítico* de $0.2 \times \sigma$, donde σ es la desviación estándar de las puntuaciones de PISA (correspondiente a $0.2 \times 75.696 = 15.139$ puntos). De este modo, cualquier variable que reporte diferencias entre valores o categorías extremas superiores a 15.139 puntos indica cambios significativos. La ventaja del uso del criterio de Cohen (también denominado efecto de tamaño) es que no se ve afectado por el tamaño de la muestra, que es el principal problema que se tiene con otras pruebas de hipótesis.

Por ejemplo, para la variable NIVEL (tabla 3.4) la diferencia media entre estudiantes de Educación media superior y Secundaria es de 44.97 puntos (como se presentó en la figura 3.4); este valor dividido entre 15.139 es $d^* = 2.96$ que, de acuerdo con el criterio de Cohen, corresponde a la categoría de diferencia *mediana* y representa casi tres veces el valor crítico.

En este modelo de la situación del estudiante, el coeficiente d^* es mayor o igual que 1 en las variables analizadas, con diferencias d^* significativas (pequeña o mediana) en la puntuación de Ciencias (tabla 3.7), con excepción de la variable REZAGO que se encuentra en el límite con 0.99. Las variables en conjunto generan una diferencia que cae en la categoría de *grande*.

Tabla 3.7 Valores de d^* para las variables del modelo situación

Variable explicativa	Coeficiente	Categoría para resultado inferior	Categoría resultado superior	Diferencia γ Variable	d^*	Significancia y clasificación
NIVEL	44.97	Secundaria = 0	Educación media superior = 1	44.97	2.96	Mediana
SC02Q01	29.32	Público = 0	Privado = 1	29.32	1.93	Pequeña
URBANORURAL	32.96	Rural = 0	Urbano = 1	32.96	2.17	Mediana
GENERO	-15.86	Mujer = 1	Hombre = 0	15.86	-1.05	Pequeña
REZAGO	-14.95	Rezagado = 1	Regular = 0	14.95	-0.99	Muy pequeña
Total teórico (valor absoluto)				138.06	9.10	Grande
Equivalente en desviaciones estándar de Ciencias				1.82		
Equivalente respecto de la escala de puntuaciones en Ciencias %				23.01		

Los valores de diferencia, así como los del coeficiente d^* son aditivos. Por ejemplo, el hecho de que un estudiante esté rezagado ($REZAGO=1$) y además se encuentre en Secundaria ($NIVEL=0$) le produce un doble efecto, porque las dos características le ocasionan una reducción en el desempeño teórico esperado. Esta combinación aditiva para d^* proporciona un valor de 3.96 (resultado de $2.96+1.00$), es decir, casi cuatro veces la diferencia crítica en su desempeño en Ciencias respecto a otro estudiante de Educación media superior; esta diferencia pertenece prácticamente a la categoría de *grande*.

A partir de estos valores, se puede reconocer que la peor situación la tienen las mujeres rezagadas, correspondientes a escuelas secundarias públicas rurales y la mejor situación ocurre con los estudiantes varones, de media superior en escuelas privadas urbanas. La diferencia entre estas situaciones extremas es de 138.06 puntos de la escala PISA, que representa una diferencia *grande* de 9.1 veces el valor crítico de diferencia para el criterio de Cohen; es decir, 1.82 veces la desviación estándar total de las puntuaciones de PISA ($\sigma = 75.696$), equivalente a 23% de la escala total de PISA.

3.3.7 Conclusiones e implicaciones de los resultados

El análisis multinivel de las variables consideradas en la situación del estudiante muestra que son muy importantes para explicar el desempeño en Ciencias dentro de la prueba PISA. A pesar de ser variables de situación es factible establecer políticas de atención para cada una de ellas, incluyendo la diversidad de género y el rezago.

En particular, se debe resaltar que para este estudio el género no se considera simplemente como una condición personal determinada al nacimiento, sino que sobre la persona inciden aspectos socioculturales asociados con el hecho de ser hombre o mujer. Visto sólo como cualidad de género podría parecer algo inmutable de la persona y la variable serviría solamente para clasificar a las personas, pero al verlo como atributo sociocultural que interactúa con el individuo, resulta de interés analizar por qué el género influye en los resultados de los estudiantes en PISA. Se trata, pues, de

una variable que debe atenderse por medio de mediaciones psicopedagógicas en el aula y en la escuela, apoyadas con políticas de interés familiar y nacional.

La situación del estudiante con rezago se presta a una doble interpretación: por una parte se puede pensar que el simple hecho de que un alumno pase a educación media superior en lugar de permanecer rezagado en secundaria, le proporciona una esperanza de puntuación más alta en la prueba PISA; esta interpretación se justifica porque el nivel educativo le proporciona un entorno más favorable para su desarrollo académico, motivacional y de autoestima. Otra forma de interpretar el resultado sería afirmando que las puntuaciones bajas que obtienen los estudiantes rezagados son producto de sus mismas deficiencias, lo cual respaldaría la decisión de no promoverlos hacia el nivel siguiente y, en su lugar, considerar una mediación psicopedagógica acorde con su nivel. ¿Cuál de estas dos posiciones puede ser más razonable? Obsérvese que la diferencia que reporta el modelo entre estudiantes de secundaria y de educación media superior es, por sí misma, casi tres veces más grande que el hecho de estar rezagado, esto apunta a que el esfuerzo que debe hacer el estudiante rezagado por alcanzar a sus compañeros en el nivel siguiente es mucho menor si pasa al nivel medio superior que si permanece en secundaria, porque el efecto se hace aditivo.

La comparación por niveles refleja un mejor desempeño de los estudiantes de educación media superior respecto de los de secundaria. Esto no es una deficiencia si se concibe que el nivel académico incide en una ganancia en las competencias de los estudiantes; en cambio si se espera que el nivel escolar no debe tener influencia en el desempeño de los estudiantes de 15 años en sus competencias para la vida, asumiendo que son adquiridas de forma no necesariamente académica, entonces se tiene un motivo de revisión del sistema educativo para propiciar el desarrollo de las competencias desde el nivel de Secundaria.

Si se considera solamente al conjunto de estudiantes de educación media superior, de escuelas privadas urbanas, su puntuación media teórica esperada sería de 478.82 puntos en la escala de PISA (es la suma de 371.57, 44.97, 29.32 y 32.96). El resultado ubica a este subgrupo de estudiantes en posición similar a los de Italia, Portugal o Grecia en los resultados generales de PISA y por arriba de 50 puntos de la posición inicial entre países comparables (ver Díaz, Flores y Martínez, 2007).

Sugerencias

Con los valores obtenidos de este primer análisis, se pueden sugerir varias acciones para mejorar la situación del estudiante, por ejemplo:

1. Incrementar el nivel en competencias científicas de las mujeres y enfatizar su papel en el desarrollo de esta rama del conocimiento, para contrarrestar su menor desempeño mostrado en Ciencias. Algunas de las acciones que se pueden tomar son: promover entre ellas el interés y los valores ambientales o ecológicos, además de realizar prácticas de solución de problemas científicos.
2. Motivar y reforzar el aprendizaje de los estudiantes rezagados ubicados en secundaria, integrando en las asignaturas actividades que favorezcan la compren-

sión de los fenómenos reales y le proporcionen herramientas de apoyo al análisis de problemas. En este nivel se debe trabajar con aspectos motivacionales, de autoestima y otros atributos que propicien el incremento del desempeño de los estudiantes rezagados en secundaria.

3. En el caso de estudiantes deficientes (potencialmente rezagados) y que hayan pasado al nivel de educación media superior, establecer esquemas de recuperación que pueden resultar ser mucho más *ligeros* y económicos que el largo proceso de repetir todo el último año de secundaria.
4. En estudios a futuro, se deberán contrastar los resultados que tiene México, por nivel escolar, con los que se tienen en otros países, para dimensionar si la diferencia que se obtiene de 44.97 puntos en la escala PISA implica una ganancia académica o no.
5. Propiciar el desarrollo de competencias para la vida en el nivel de secundaria, en particular las asociadas con la comprensión de los fenómenos y su relación con las ciencias, la forma de resolver problemas y el modo de razonar en esta área de conocimiento.
6. Reforzar las escuelas rurales para que reduzcan la brecha con las escuelas urbanas. En este caso puede tratarse de políticas de la institución con posible apoyo estatal.
7. En un estudio de validez facial, de contenido y de constructo, revisar detalladamente el contexto de los reactivos de PISA para identificar si favorecen al medio ambiente urbano respecto del rural, especialmente si se toman en cuenta las diferencias estructurales que se tienen en el medio rural mexicano contra las condiciones de otros países. Esto propiciaría un estudio comparativo entre países de Latinoamérica para contrastar los dos tipos de ambientes y establecer similitudes o diferencias contra los resultados obtenidos por los estudiantes de México.

3.4 Modelos de la dimensión de equidad

3.4.1 Pregunta de investigación

Se pueden plantear varias preguntas en relación con la equidad en el sistema educativo ante la prueba PISA:

- ¿Cómo interviene el índice socioeconómico como fuente de inequidad en el desempeño de los estudiantes ante la prueba PISA?
- ¿Interviene el estatus socioeconómico de la escuela en el desempeño estudiantil?
- Además del estatus socioeconómico, ¿existen variables que puedan englobarse en la dimensión de equidad y que participen, de manera estadísticamente significativa, en el desempeño de los estudiantes?

3.4.2 Hipótesis

Las hipótesis en relación con la equidad son:

- Las diferencias de desempeño de los estudiantes ante la prueba PISA están asociadas con una inequidad debida al estatus socioeconómico.
- El estatus socioeconómico de la escuela es un factor estadísticamente significativo en el desempeño de los estudiantes en la prueba PISA.
- La escolaridad y el nivel educativo de los padres inciden en las diferencias en el desempeño de los estudiantes de manera estadísticamente significativa.
- Hay características organizacionales de la escuela que pueden producir diferencias estadísticamente significativas en el desempeño en Ciencias de los estudiantes.

3.4.3 Variables explicativas

Por medio del análisis factorial exploratorio en cada nivel de anidamiento se identifican dos posibles modelos. El componente factorial del nivel 1, asociado con la variable EQUIDAD, contiene al índice socioeconómico ESCS junto con otras variables que forman parte del mismo ESCS. Este grupo de variables se analizó por separado para confirmar su interdependencia y, con ello, simplificar el modelo utilizando solamente el ESCS, con lo que resultó un modelo más parsimonioso en el nivel 1. En el nivel 2 se incluyeron variables relacionadas con la organización de la escuela, asociadas con el componente factorial que se interpretó también como EQUIDAD.

Tabla 3.8 Variables del modelo de equidad

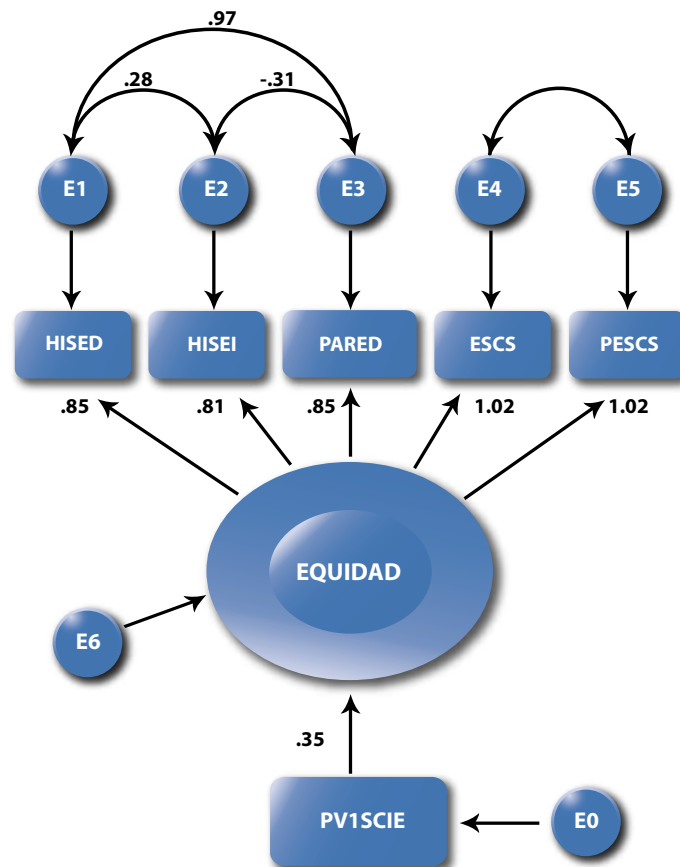
Nivel	Variables	Descripción
Variables de nivel 1	ESCS	Índice de estatus socioeconómico del estudiante
	HISEI	Ocupación de mayor jerarquía entre los padres del estudiante
	HISCED	Grado académico de mayor jerarquía entre los padres del estudiante
	PARED	Grado académico de mayor jerarquía entre los padres del estudiante medido en años escolares
	PESCS	Ubicación decilar del ESCS del estudiante
Variables de nivel 2	XESCS	Media del ESCS a nivel escuela
	SC02Q01	Tipo de sostenimiento de la escuela: 0= Público, 1=Privado
	SCMATEDU	Calidad de los recursos educativos de la escuela
	IRATCOMP	Porcentaje de computadoras para trabajo académico por estudiante
	IRRESPCU	Autonomía de la escuela para establecer y evaluar el <i>currículum</i>
	NIVEL	Nivel de estudios que imparte la escuela 0= Secundaria 1= Educación media superior
	URBANORURAL	Tipo de comunidad en la que se encuentra la escuela: 0=Rural, 1=Urbano

3.4.4 Resultados del análisis factorial confirmatorio

Modelo de equidad en el nivel I

La variable mediadora EQUIDAD en el nivel I, tiene un coeficiente de regresión estandarizada muy similar entre las variables consideradas (de 0.81 a 1.02) y un coeficiente de regresión estructural de más de un tercio de (0.35) con el desempeño en Ciencias, apreciándose un ajuste alto (AGFI=0.965).

Figura 3.5 Modelo estructural de la dimensión de equidad en el nivel I



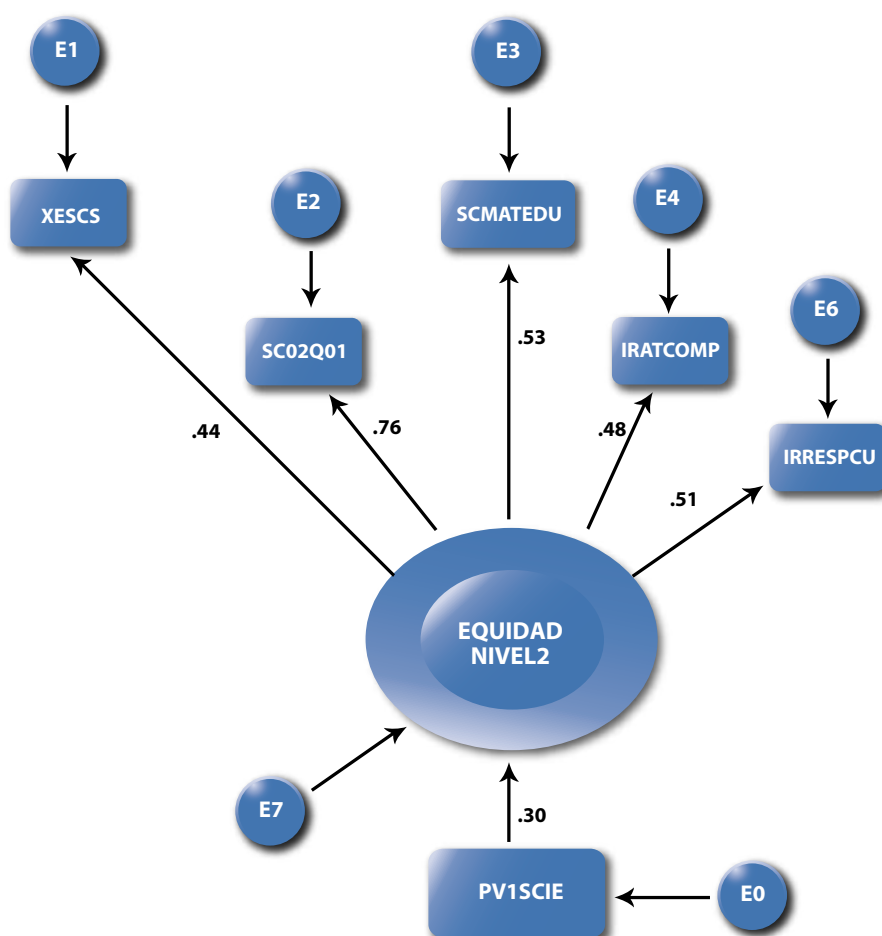
AGFI=0.965

Es claro que ESCS y PESCS deben tener prácticamente la misma contribución (1.02) por la alta correlación entre ellas; el modelo resulta idéntico si se modela esta correlación o no.

Modelo de equidad en el nivel 2

Al construir una variable latente mediadora EQUIDAD en el nivel 2 se obtiene un buen ajuste ($AGFI=0.893$) aunque ligeramente menor que en el nivel 1. El modelo indica que el índice socioeconómico a nivel de la escuela (XESCS) se relaciona con un coeficiente de regresión estandarizado de 0.44, el menor de todo el conjunto, teniendo un coeficiente más alto para el tipo de escuela (SC02Q01). Otras variables de organización de la escuela como la proporción de computadoras para uso académico (IRATCOMP) y la responsabilidad del *currículo* (IRRESPCU) tienen valores similares entre sí (0.48 a 0.53).

Figura 3.6 Modelo estructural de la dimensión de equidad en el nivel 2



AGFI=0.893

Al parecer, el conjunto tiene una influencia del orden de 0.30 en la puntuación de Ciencias.

Análisis de las variables correlacionadas y modelado del ESCS

Al revisar el modelo de equidad del nivel I se aprecia una gran correlación entre las variables, debido a que el diseño del ESCS contiene a las variables HISEI, HISCED y PARED que resultaron dentro del mismo componente factorial. Se elaboró un modelo adicional para verificar la combinación de estas variables en la predicción del ESCS, para evitar incluir variables altamente correlacionadas entre sí en el mismo modelo. Con este antecedente, el modelo para las variables que participan en la dimensión de equidad se plantea en dos etapas:

1. Identificar la relación multinivel que se tiene entre las variables explicativas y el índice ESCS propuesto por el proyecto PISA.
2. Establecer el modelo multinivel para predecir el resultado en Ciencias a partir de la variable mediadora ESCS confirmada en la etapa A.

El análisis multinivel de la primera etapa es un caso donde la variable de salida a modelar es el ESCS y se incluyen las variables predictoras restantes identificadas como parte de la dimensión de equidad. Un procedimiento para identificar al conjunto de variables involucradas en esta dimensión podría realizarse por aproximaciones sucesivas, combinando las variables hasta encontrar el modelo de mayor ajuste; este procedimiento no es seguido aquí. El segundo método parte de las capacidades de análisis exploratorio disponible en el *software* HLM para los niveles 2 y 3; se trata de un análisis sistemático realizado por el mismo *software* y con el cual se reconocen las variables que pueden ser significativamente predictoras de la variable dependiente ESCS. Con esta técnica se concluye que no hay variables de niveles 2 y 3, por lo que el modelo se construye sólo con variables del nivel I.

$$\text{ESCS}_{ijk} = \gamma_{000} + \gamma_{100} \text{HISEI}_{ijk} + \gamma_{200} \text{BAHISEI}_{ijk} + \gamma_{300} \text{HISCED}_{ijk} + \gamma_{400} \text{BAHISCED}_{ijk} + \gamma_{500} \text{PARED}_{ijk} + r_{0jk} + u_{00k} + e_{ijk}$$

Una vez identificadas las variables se procede a efectuar el análisis multinivel con el modelo indicado para ESCS, obteniéndose los resultados de la tabla 3.9.

Tabla 3.9 Resultados del modelo multinivel del índice socioeconómico ESCS

Efecto fijo		Coefficiente	Error estándar	t	Grados de libertad	p-valor
Interceptos						
Valor base de ESCS	γ_{000}	-3.915	0.07	-52.77	6	0.000
Pendientes						
HISEI	γ_{100}	0.031	0.00	157.68	30345	0.000
BAHISEI	γ_{200}	-0.190	0.04	-4.28	30345	0.000
PARED	γ_{300}	0.164	0.00	76.63	30345	0.000
BAPARED	γ_{400}	0.177	0.13	1.40	30345	0.161

Efecto aleatorio	Nivel	Unidad		Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
1	Estudiante	e_{ijk}		0.34	0.11				67.2
2	Escuela	r_{0jk}		0.21	0.04	1081	9908.80	0.000	25.2
3	Región	u_{00k}		0.11	0.01	6	397.89	0.000	7.6

Desajuste (Promedio)	23037.20
----------------------	----------

Por los resultados obtenidos, se hace evidente que la varianza del ESCS todavía puede ser explicada en un 67.2% por variables predictoras del primer nivel. Las variables disponibles tienen una participación significativa, con excepción de la variable muda que se asignó para los casos de omisión, en tanto que el HISCED pudo eliminarse del modelo, lo cual es justificable porque ya se encuentra incluido en el diseño de la variable PARED.

El desajuste de las variables para el modelo del ESCS es pequeño, tan sólo de 23 mil 037 (este valor no debe compararse contra los modelos anteriores relativos al desempeño en Ciencias por tratarse de una variable dependiente distinta).

Con esta construcción se simplifica el modelo de equidad, al intervenir solamente al ESCS porque engloba al resto de variables explicativas con las cuales se construyó. El resto del modelo de equidad se apoya en las variables que se descartaron por no estar significativamente asociadas con el ESCS (descartadas en el modelo exploratorio por el propio *software*).

3.4.5 Ecuaciones de los modelos multinivel

En este modelo se tomaron en cuenta las interacciones entre las variables utilizadas en el proyecto PISA para definir los índices, comentadas en la sección previa. Ahora se construyeron dos modelos: el primero utiliza el índice socioeconómico ESCS incluido en la base de datos de PISA y su variable *bandera*; el segundo, parte de lo indicado en algunas referencias (por ejemplo Willms, 2006), donde se señala el interés de modelar la no linealidad respecto al índice socioeconómico considerando un término cuadrático (ESCS2) y también tres variables que habrían participado como parte de la variable SITUACION en el nivel 1, pero que ahora se refieren al nivel 2: NIVEL, SC02Q01 e URBANORURAL. Las ecuaciones de estos dos modelos son las siguientes:

$$PVSCIE_{ijk} = \gamma_{000} + \gamma_{010} XESCS_{jk} + \gamma_{020} NIVEL_{jk} + \gamma_{030} SC02Q01_{jk} + \gamma_{040} IRRESPCU_{jk} + \gamma_{100} ESCS_{ijk} + \gamma_{200} BAESCS_{ijk} + r_{0jk} + u_{00k} + e_{ijk}$$

$$PVSCIE_{ijk} = \gamma_{000} + \gamma_{010} SCMATEDU_{jk} + \gamma_{020} NIVEL_{jk} + \gamma_{030} URBANORURAL_{jk} + \gamma_{100} ESCS_{ijk} + \gamma_{200} BAESCS_{ijk} + \gamma_{300} ESCS2_{ijk} + r_{0jk} + u_{00k} + e_{ijk}$$

Ambos modelos consideran los valores plausibles y la ponderación regional de las variables del nivel 1 y sus variables cuentan con coeficientes diferentes de cero en forma estadísticamente significativa.

Tabla 3.10 Resultados del primer modelo de equidad con ESCS

Efecto fijo		Coeficiente	Error estándar	t	Grados de libertad	p-valor
Interceptos						
Puntuación base en Ciencias	γ_{000}	443.832	16.21	27.39	6	0.000
XESCS	γ_{010}	38.741	9.02	4.30	1083	0.000
NIVEL	γ_{020}	35.699	7.74	4.61	1083	0.000
SC02Q01	γ_{030}	-31.154	11.50	-2.71	1083	0.007
IRRESPCU	γ_{040}	9.981	2.21	4.52	114	0.000
ESCS	γ_{100}	6.355	0.77	8.21	9	0.000
BAESCS	γ_{200}	-38.573	18.09	-2.13	30	0.041

Efecto aleatorio			Desviación estándar	Componente de varianza	Grados de libertad	ESCS ²	p-valor	Porcentaje de varianza
Nivel	Unidad							
1	Estudiante	e_{ijk}	58.16	3383.05				76.1
2	Escuela	r_{ojk}	32.26	1040.59	1077	8401.12	0.000	23.4
3	Región	u_{ook}	4.97	24.70	6	29.16	0.000	0.6

No.	Nivel		Componente de varianza en el modelo nulo	Componente de varianza en este modelo	Diferencia de varianza %
1	Estudiante	e_{ijk}	3416.98	3383.05	1.0
2	Escuela	r_{ojk}	3063.17	1040.59	66.0
3	Región	u_{ook}	441.09	24.70	94.4

Desajuste (Promedio)	335330.63	Mejora respecto del modelo de referencia	0.40 %	
Diferencia respecto del modelo nulo	1334.50	Diferencia en grados de libertad	6	Diferencia significativa

Este modelo desajusta menos que el modelo incondicional, con una diferencia estadísticamente significativa, representando una mejora de 0.4% que se traduce en un modelo que explica 94.4% de la varianza del nivel 3 y 66% de la varianza en el nivel 2, aunque no incide en variables explicativas para la varianza en el nivel 1.

El modelo donde se incluye el término cuadrático del estatus socioeconómico representa una mejora más pequeña que el caso anterior al tenerse sólo 0.29% de reducción en el desajuste (tabla 3.11). Aunque se trata de un modelo menos ajustado que el primero de equidad, se tiene el interés de analizar el efecto de cada variable en el desempeño en Ciencias.

Tabla 3.11 Resultados del modelo de equidad con el término cuadrático

Efecto fijo		Coeficiente	Error estándar	t	Grados de libertad	p-valor
Interceptos						
Puntuación base en Ciencias	γ_{000}	379.015	13.08	28.97	6	0.000
SCMATEDU	γ_{010}	7.770	2.61	2.97	1084	0.003
NIVEL	γ_{020}	46.502	5.84	7.96	1084	0.000
URBANORURAL	γ_{030}	26.852	6.96	3.86	386	0.000
Pendientes						
ESCS	γ_{100}	9.393	0.76	12.33	18	0.000
BAESCS	γ_{200}	-36.560	18.44	-1.98	31	0.056
ESCS2	γ_{300}	1.389	0.43	3.24	29	0.003

Efecto aleatorio			Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
Nivel	Unidad							
1	Estudiante	e_{ijk}	58.10	3375.72				65.5
2	Escuela	r_{ojk}	40.59	1647.70	1078	11426.89	0.000	32.0
3	Región	u_{00k}	11.25	126.66	6	75.37	0.000	2.5

No.	nivel		Componente de varianza en el modelo nulo	Componente de varianza en este modelo	Diferencia de varianza %
1	Estudiante	e_{ijk}	3416.98	3375.72	1.2
2	Escuela	r_{ojk}	3063.17	1647.70	46.2
3	Región	u_{00k}	441.09	126.66	71.3

Desajuste (Promedio)	335698.27	Mejora respecto del modelo de referencia	0.29%	
Diferencia respecto del modelo nulo	966.86	Diferencia en grados de libertad	6	Diferencia significativa

3.4.6 Presentación e interpretación de resultados

Las diferencias entre valores extremos de las variables son de gran interés, en particular el nivel socioeconómico de la escuela (XESCS) tiene un efecto grande ($d^* = 12.28$) en el primer modelo y se revela como la variable más notable de todas. De este modo, se tiene que una escuela privada de educación media superior con el nivel socioeconómico más alto, donde los directivos están involucrados en las decisiones sobre el currículo, genera la máxima diferencia de 366 puntos de la escala PISA, respecto de otra escuela secundaria con nivel socioeconómico bajo y con el personal directivo de perfil bajo en relación con el currículo; esta diferencia es de 24.18 veces el valor crítico para tener diferencias significativas con el criterio de la d de Cohen.

Tabla 3.12 Valores de d^* para las variables del modelo de equidad con el ESCS

Variable explicativa	Coeficiente	Categoría para resultado inferior	Categoría resultado superior	Diferencia y Variable	d^*	Significancia y clasificación
XESCS	38.741	-3.13	1.67	185.96	12.28	Grande
NIVEL	35.699	Secundaria = 0	Educación media superior = 1	35.70	2.36	Pequeña
SC02Q01	-31.154	Pública = 1	Privado = 0	31.15	-2.06	Pequeña
IRRESPCU	9.981	-1.4	1.27	26.65	1.76	Pequeña
ESCS	6.355	-4.83	2.72	47.98	3.17	Mediana
BAESCS	-38.573	Imputado = 1	No imputado = 0	38.57	-2.55	Mediana
Total teórico (valor absoluto)				366.01	24.18	Grande
Equivalente en desviaciones estándar de Ciencias				4.8353		
Equivalente respecto de la escala de puntuaciones en Ciencias %				61.002		

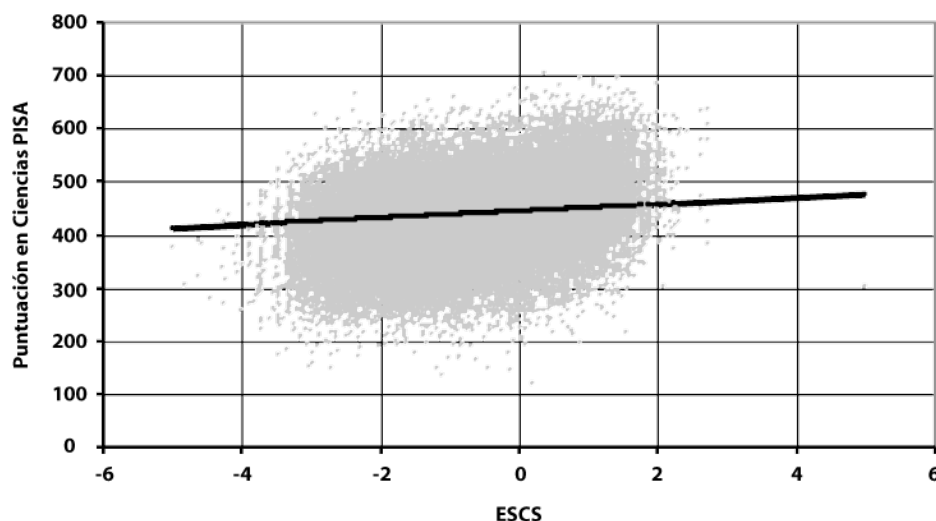
Cuando se considera el término de segundo grado en ESCS2, la variable que reporta mayor efecto es el propio estatus socioeconómico del estudiante, pero no puede compararse directamente con el modelo anterior, por la presencia del término cuadrático. La diferencia entre los casos extremos llega a 256.5 puntos en la escala PISA.

Tabla 3.13 Valores de d^* para las variables del modelo de equidad con el ESCS2

Variable explicativa	Coeficiente	Categoría para resultado inferior	Categoría resultado superior	Diferencia y Variable	d^*	Significancia y clasificación
SCMATEDU	7.770	-3.43	2.14	43.28	2.86	Mediana
NIVEL	46.502	Secundaria = 0	Educación media superior = 1	46.50	3.07	Mediana
URBANORURAL	26.852	Rural = 0	Urbano = 1	26.85	1.77	Pequeña
ESCS	9.393	-4.83	2.72	70.92	4.68	Grande
BAESCS	-36.560	Imputado = 1	No imputado = 0	36.56	-2.41	Pequeña
ESCS2	1.389	0	23.32	32.39	2.14	Pequeña
Total teórico (valor absoluto)				256.5	16.94	Grande
Equivalente en desviaciones estándar de Ciencias				3.39		
Equivalente respecto de la escala de puntuaciones en Ciencias %				42.75		

Es conveniente comparar los dos modelos, cuando menos en la parte correspondiente al estatus socioeconómico de primero y de segundo grado. Para ello se representa el diagrama de dispersión considerando el ESCS en el eje X y el primer valor plausible de PISA en el eje Y. La recta trazada corresponde con el modelo de primer grado en ESCS que se acaba de obtener:

Figura 3.7 Diagrama de dispersión ESCS vs. Puntuación de Ciencias, modelo de primer grado



Resulta interesante comparar el modelo multinivel con el modelo de regresión lineal que se obtiene con ESCS como variable predictora y el primer valor plausible en Ciencias de cada estudiante, pero sin considerar el anidamiento ni la ponderación o peso por cada registro.

El modelo de regresión simple se define de esta manera:

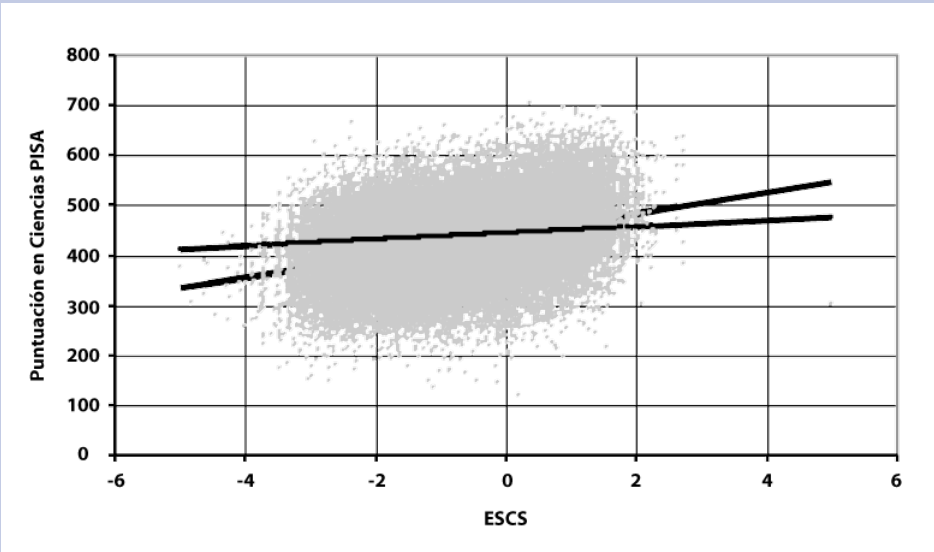
$$PVSCIE_{ijk} = a + b \text{ ESCS}_{ijk} + e$$

Obteniéndose:

$$PVSCIE_{ijk} = 440.22 + 21.416 \text{ ESCS}_{ijk} + e$$

Con un coeficiente de determinación $r^2=0.1273$, y una correlación $r = 0.36$.

Figura 3.8 Comparación del modelo multinivel y el modelo de regresión lineal del ESCS



Esta misma combinación de variables, pero en un modelo multinivel considerando los anidamientos por escuela y por región, conduce a:

$$PVSCIE_{ijk} = 423.63 + 7.71 \text{ ESCS}_{ijk} + r_{ojk} + u_{00k} + e_{ijk}$$

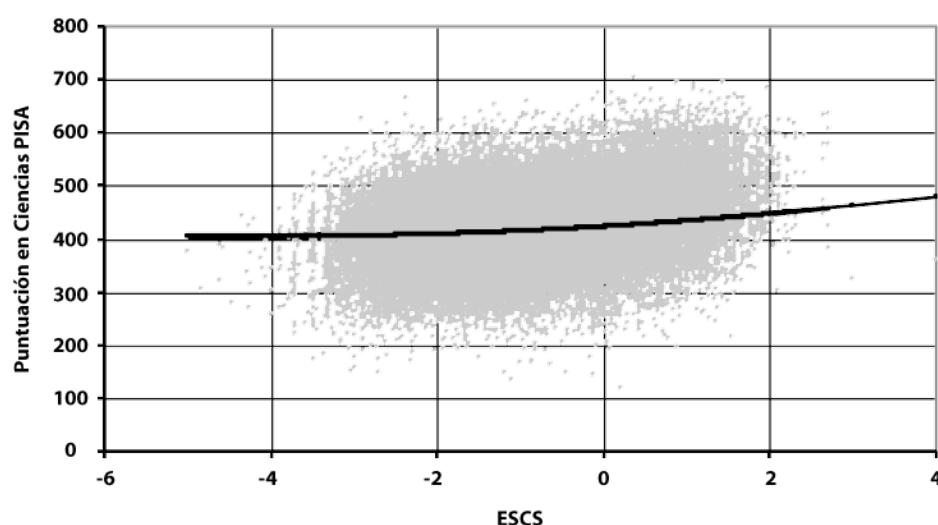
Efecto aleatorio			Desviación estándar	Componente de varianza	Porcentaje de varianza
nivel	Unidad				
1	Estudiante	e_{ijk}	58.166	3383.323	53.7
2	Escuela	r_{ojk}	50.963	2597.277	41.2
3	Región	u_{00k}	17.839	318.245	5.1

Como puede observarse, el intercepto es muy similar entre los dos modelos (la diferencia entre ambos es del 3.7 %), pero la pendiente sí es significativamente distinta, teniéndose una sobreestimación de 64% al usar una regresión simple (que no considera anidamiento).

La ganancia que se obtiene con el modelo de segundo grado en ESCS parece no ser significativa respecto del modelo de primer grado, pero obsérvese la representación de los puntos en el plano ESCS vs. puntuación en Ciencias se presenta en la figura 3.9.

$$PVSCIE_{ijk} = 422.25 + 9.457 \text{ ESCS}_{ijk} + 1.252 \text{ ESCS}_{ijk}^2 + r_{0jk} + u_{00k} + e_{ijk}$$

Figura 3.9 Diagrama de dispersión ESCS vs. Puntuación de Ciencias, modelo de segundo grado



La forma que adopta el modelo, parece combinar parcialmente los dos modelos lineales (la pendiente del multinivel para la parte superior del estatus socioeconómico ESCS y la de la regresión lineal para los valores bajos del ESCS), lo cual da elementos para preferirlo respecto a uno solo de los modelos lineales vistos por separado. Ahora se puede calcular la diferencia de desempeño que se tiene entre los valores extremos de ESCS (de -4.83 a +2.93), alcanzando 51.46 puntos de la escala PISA, que representa 67.99% de la desviación estándar de puntuaciones obtenidas en México.

Puede demostrarse que este modelo cuadrático tiene un mínimo de desempeño modelado en Ciencias de 404.40 puntos en la escala PISA, para $\text{ESCS} = -3.777$, valor que está prácticamente fuera del rango de valores de este índice⁹; debe tomarse en cuenta que el modelo puede carecer de sentido para predecir el desempeño de personas por debajo de tal índice socioeconómico.

⁹ Este valor se obtiene derivando la expresión cuadrática $Y=a+bx+cx^2$, obteniéndose el mínimo para $x=-b/(2c)$. La verificación del mínimo en modelos cuadráticos no se acostumbra hacer en muchos análisis multinivel, dando lugar a interpretaciones inaceptables, especialmente en los valores inferiores de la variable explicativa.

3.4.7 Conclusiones e implicaciones de los resultados

En este momento, resulta importante emitir un juicio de valor sobre la equidad representada por los modelos propuestos. La interpretación más trivial de la equidad consiste en esperar que todos los estudiantes, independientemente de su estatus socioeconómico y cultural, obtengan puntuaciones similares en la prueba PISA; esto obliga a que el modelo lineal tenga una pendiente nula (coeficiente nulo en ESCS). Es evidente que obtener un valor nulo es imposible en la práctica y, por lo tanto, debe considerarse que es nulo dentro del error de diseño del modelo, lo cual transforma el problema conceptual y cualitativo en un análisis de significancia estadística para los coeficientes del modelo multinivel o, como se plantea aquí, utilizando el criterio de Cohen que reporta el efecto de tamaño sin verse afectado por el número de sujetos en la muestra (ver anexo A2.8).

Se puede aplicar el modelo multinivel para determinar los valores de ESCS donde ocurre el cambio o diferencia unitaria para d^* (tabla 3.14).

Tabla 3.14 Valores de ESCS para cambios unitarios en d^*

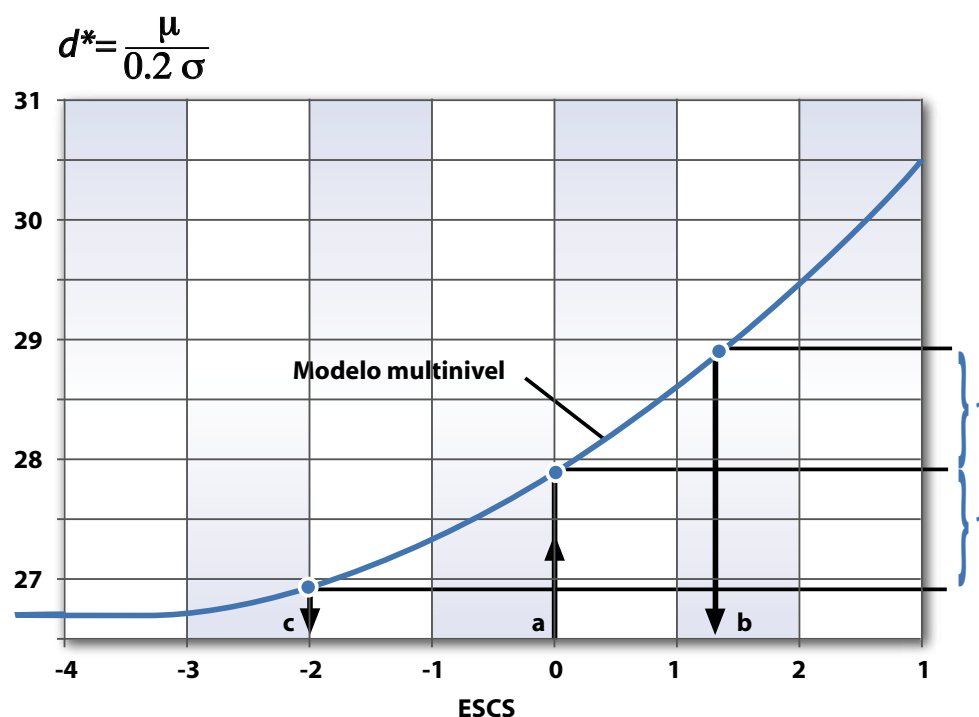
ESCS	PVTeórica	$\frac{\mu}{0.2\sigma}$	$d^* = \frac{\mu_a - \mu_b}{0.2\sigma}$	Diferencia en ESCS	Diferencia en desviaciones estándar de ESCS
-5.000	406.27	26.84			
-0.091	421.40	27.84	1.00	4.909	3.90
1.291	436.54	28.84	1.00	1.381	1.10
2.369	451.68	29.84	1.00	1.078	0.86
3.284	466.81	30.83	1.00	0.915	0.73

Puede utilizarse la tabla para localizar los puntos en los cuales ocurre un cambio cualitativo entre dos puntuaciones de Ciencias, lo cual se asocia con una posible inequidad debida al estatus socioeconómico. Si se empieza con el valor más bajo del ESCS, se tiene que para los niveles inferiores se requiere de un cambio de 4.909 en ESCS (casi cuatro desviaciones estándar del propio ESCS); pero si se empieza por el valor más alto, entonces sólo se requiere de un cambio de 0.915 en ESCS (menos de 73% de la desviación estándar). Los cambios son más notables en la parte superior que en la inferior del ESCS.

La figura 3.10 ilustra la aplicación del modelo multinivel y del criterio de Cohen para establecer intervalos de inequidad. El eje de las abscisas corresponde al nivel socioeconómico ESCS, el eje vertical es el valor de la puntuación teórica que proporciona el modelo multinivel dividido por el valor del criterio 0.2σ . Como un ejemplo del uso de la figura, tómese el punto a en $ESCS_a=0$, con el cual se determina el punto sobre la curva del modelo multinivel de 27.9. A partir de este punto, cualquier puntuación que difiera de una unidad representa una diferencia significativa, en este caso cuando un estudiante tiene por lo menos 28.9 puntos o cuando más 26.9. Estas

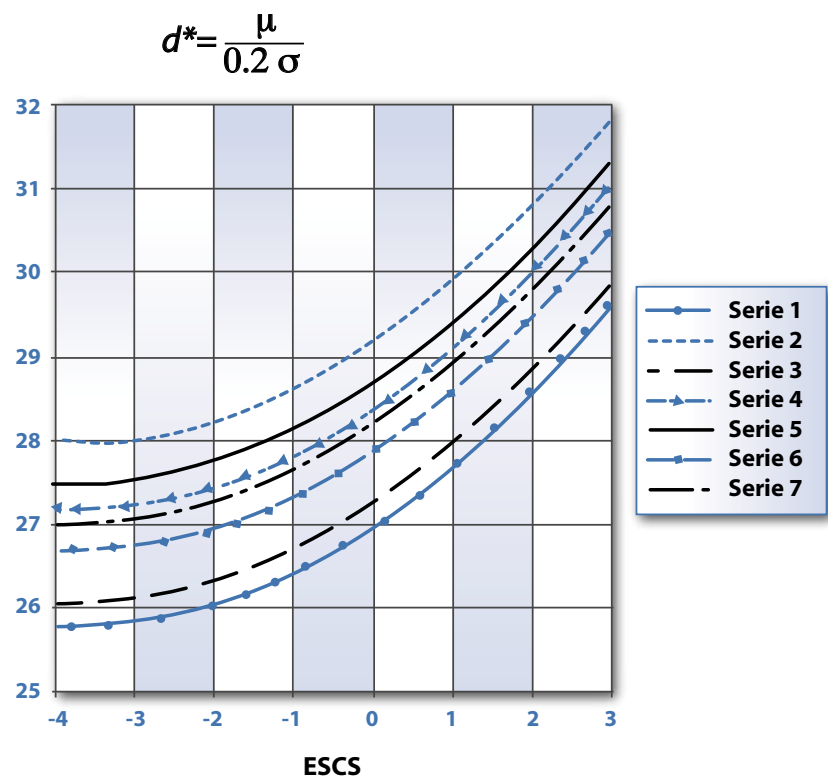
diferencias ocurren, en el caso superior en $ESCS_b = 1.36$, y en el caso inferior en $ESCS_c = -2.30$. Puede apreciarse el efecto de la no-linealidad, teniéndose una diferencia de estatus socioeconómico de 2.30 puntos al pasar a la parte baja de ESCS y de 1.36 al pasar a la parte alta de ESCS; la inequidad ocurre más rápidamente con los estratos superiores que con los inferiores. Puede afirmarse que por debajo de la media del ESCS ($ESCS = -0.8$) las diferencias en los resultados de los estudiantes no se asocian con una inequidad debida al estatus socioeconómico, en cambio, por arriba de la media el nivel socioeconómico incide de forma notable para producir inequidad en los resultados de la prueba PISA.

Figura 3.10 Intervalos de inequidad a partir del modelo multinivel y del criterio de Cohen



Por la forma en que se definió el modelo multinivel, este comportamiento es idéntico en las siete regiones, pero difieren en el intercepto, como se muestra en la figura siguiente:

Figura 3.11 Diferencias por inequidad en las siete regiones



Ahora se puede determinar la diferencia entre regiones para el intercepto global que proporciona el modelo multinivel y las medias de cada región. La tabla compara la región 2 que tiene la puntuación media más alta con el resto de regiones.

Tabla 3.15 Comparación de puntajes de inequidad entre regiones

Región	Media μ	$\frac{\mu}{0.2\sigma}$	$d^* = \frac{\mu_a - \mu_b}{0.2\sigma}$
Región 2	441.050	29.133	
Región 4	433.014	28.602	0.531
Región 7	432.010	28.536	0.597
Región 5	431.861	28.526	0.607
Región 6	421.306	27.829	1.304
Región 3	411.737	27.197	1.936
Región 1	408.588	26.989	2.144

Bajo el criterio de Cohen, se tiene una inequidad entre regiones si la diferencia de interceptos o medias normalizadas por el valor crítico $0.2 \times \sigma$ es igual a 1; esta diferencia se presenta en la columna bajo d^* , con lo cual se tienen diversas combinaciones de regiones que presentan inequidad entre ellas, con una máxima diferencia de 2.144, pero es más notable la diferencia por el estatus socioeconómico dentro de cada región que la diferencia entre regiones para un mismo valor de ESCS.

En conclusión, el análisis de las variables de la dimensión de equidad permite verificar las hipótesis planteadas para el Modelo de Calidad, mostrando que las regiones extremas tienen una participación de importancia en la equidad de los desempeños, pero las diferencias entre las personas de una misma región son mucho más importantes, debido al efecto no lineal del índice socioeconómico ESCS.

Sugerencias

Pueden hacerse varias propuestas a partir de los modelos construidos.

1. Los responsables nacionales, regionales y estatales de las políticas y acciones de apoyo socioeconómico, cultural y educativo, deberán considerar la conveniencia de inyectar una gran inversión de recursos, especialmente en los niveles bajos, para poder obtener incrementos significativos en el desempeño de los estudiantes.
2. Se debe propiciar la mejora entre las personas de los niveles socioeconómicos más altos, sabiendo que en éstos el efecto es más notable con una inversión más pequeña. Lo que indica el comportamiento no lineal, es que se deben proveer más recursos en los niveles bajos que en los altos para obtener resultados similares en las puntuaciones de Ciencias.
3. Se recomienda tener cuidado de no caer en el error de pensar que para un cierto estatus socioeconómico es preferible el cambio de región, porque puede producir variaciones sustanciales en el desempeño del estudiante sin ninguna inversión estatal o regional.
4. Si se toma en cuenta que además del estatus socioeconómico individual del estudiante, también el estatus socioeconómico de la escuela interviene en los resultados, entonces puede afirmarse que el ambiente en el que se desenvuelven los estudiantes cuenta mucho en su desempeño. Se sugiere propiciar mejores condiciones de aprendizaje y de desarrollo socioeconómico de la institución, es decir, es tarea de los directivos escolares y otras autoridades promover la mejora del ambiente que ofrezca la institución, donde se aproveche el estatus mismo de los estudiantes que ingresan en ella.
5. Al ver que los resultados son mejores en educación media superior privada urbana que en secundaria pública rural, puede pensarse que se trata de aspectos que son irremediables porque integran las condiciones y el entorno en el cual se mueve cada institución. Pero hay una opción de mejora interesante con la intervención de las autoridades y del docente mismo como responsables del desarrollo del *currículum*, se trata de una veta de oportunidad nada despreciable

para incrementar el desempeño de los estudiantes, al buscar mejores esquemas de enseñanza-aprendizaje, nuevos modelos de planeación didáctica y de desarrollo curricular.

3.5 Modelos para la dimensión de eficiencia

3.5.1 Pregunta de investigación

¿Una mayor disponibilidad de bienes económicos, materiales y culturales puede incidir en el desempeño de los estudiantes en la prueba PISA?

3.5.2 Hipótesis

La combinación de diversos bienes en el hogar y la familia está asociada con el desempeño de los estudiantes en Ciencias en la prueba PISA, de manera estadísticamente significativa.

3.5.3 Variables explicativas

Las variables de los niveles 1 y 2 que intervienen en los componentes factoriales son los siguientes:

Tabla 3.16 Variables que componen el modelo de eficiencia

Nivel	Variables	Descripción
Variables de nivel 1	HOMEPOS	Índice de posesiones en el hogar
	SC02Q01	Tipo de sostenimiento de la escuela: 0=Público, 1=Privado
	HEDRES	Recursos educativos en el hogar
	WEALTH	Índice de bienestar familiar
	CULTPOSS	Posesiones culturales en el hogar
Variables de nivel 2	IRATCOMP	Porcentaje de computadoras para trabajo académico por estudiante
	CLSIZE	Promedio del tamaño de clase
	STRATIO	Estratos en los que se dividió la muestra para formar grupos de estudiantes homogéneos que permitieran reducir la varianza.
	URBANORURAL	Tipo de comunidad en la que se encuentra la escuela: 0=Rural, 1=Urbano
	IRRESPCU	Autonomía de la escuela para establecer y evaluar el currículum

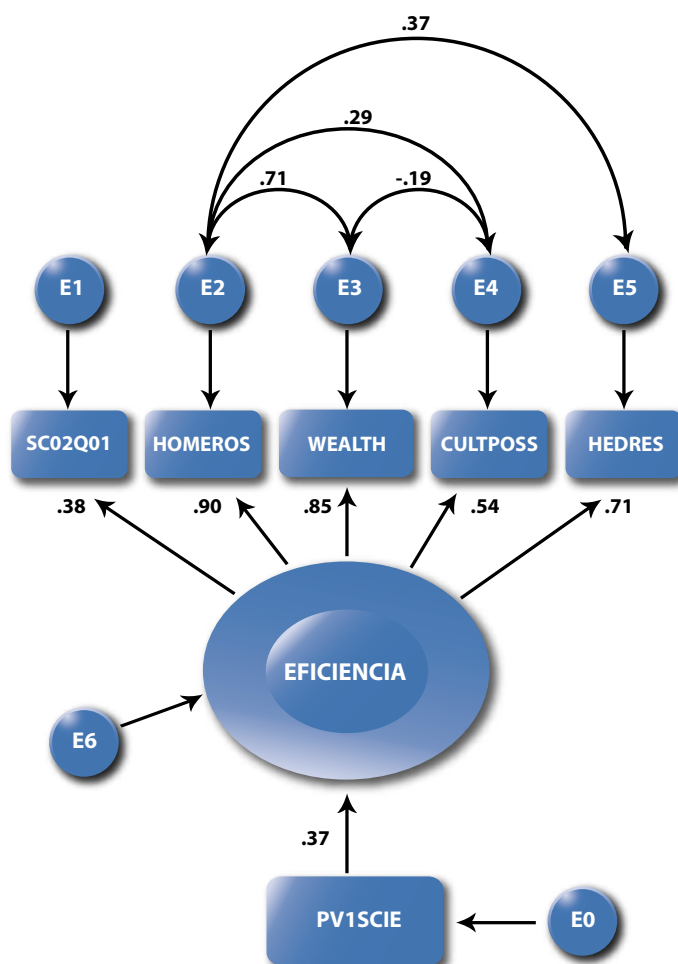


3.5.4 Resultados del análisis factorial confirmatorio

Se modelaron las variables asociándolas a una variable latente denominada EFICIENCIA con apoyo de las bases de datos de los niveles 1 y 2. Estos modelos combinan variables que implican un manejo racional y productivo de los recursos del hogar y de la familia. En el nivel escolar participan variables que se refieren a relaciones de los bienes informáticos, el tamaño de clase y la razón del número de estudiantes por docente.

El modelo propuesto para Eficiencia en el nivel 1 es un poco más complejo al tener involucradas las correlaciones entre variables, apreciándose un grupo de variables relativo a condiciones del hogar y familiares, con un ajuste muy bueno (AGFI=0.976).

Figura 3.12 Modelo estructural de la dimensión de eficiencia con variables de nivel 1

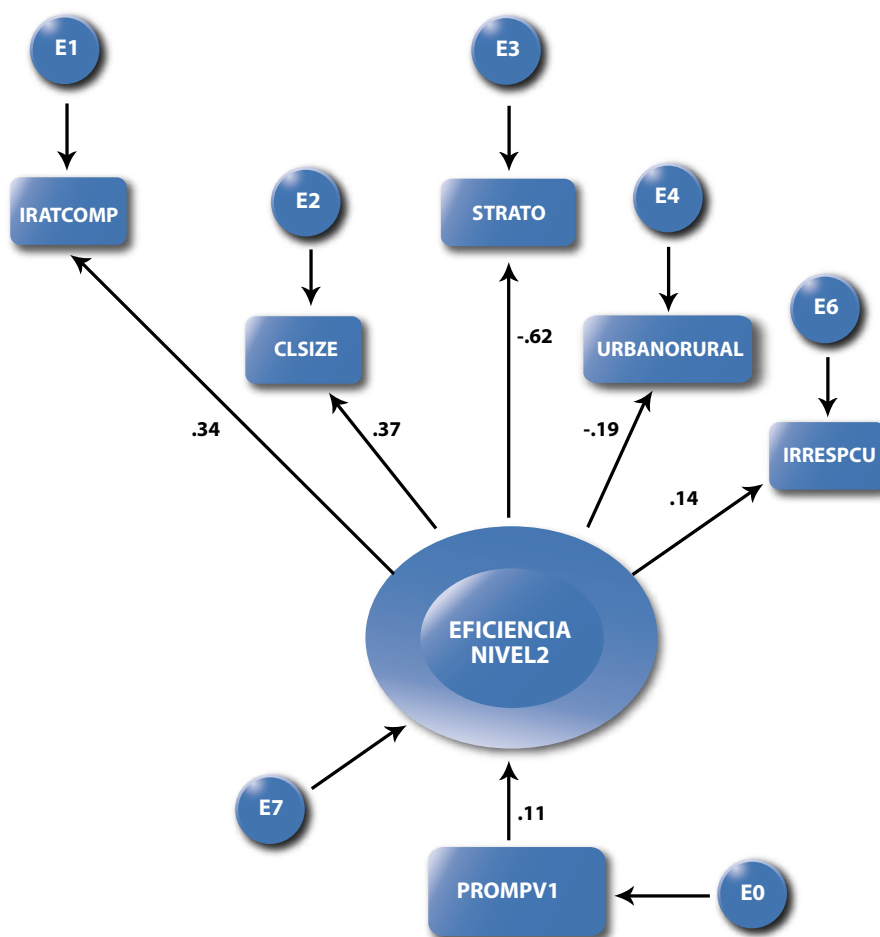


AGFI=0.976

Las variables del hogar están todas correlacionadas (la mayor correlación se ubica entre WEALTH, índice de bienestar en el hogar y HOMEPOS, índice de posesiones en el hogar) y como conjunto tienen los coeficientes más altos respecto a la variable latente. Este conjunto se relaciona con cambios en la regresión de 0.37 para el desempeño en Ciencias.

La segunda organización de la variable latente EFICIENCIA en el nivel 2 agrega los índices IRATCOMP (relación de computadoras para uso académico), IRRESPCU (índice de responsabilidad del currículum en la escuela) y la relación docente-estudiante, STRATIO, que produce un efecto negativo inesperado en la variable latente. Este modelo estructural es aceptable (AGFI=0.905).

Figura 3.13 Modelo estructural de la dimensión de eficiencia con variables de nivel 2



AGFI=0.905

Modelado de HOMEPOS

Debido a la correlación entre variables del nivel I, para este modelo se consideró necesario hacer un análisis de la asociación de las variables que conforman al índice de posesiones del hogar HOMEPOS (que incluye las respuestas de las preguntas correspondientes a WEALTH, CULTPOSS Y HEDRES), para llegar a establecer la ecuación multinivel más parsimoniosa, con menos variables explicativas.

Se plantea el modelo de HOMEPOS de manera similar a lo que se realizó con el índice de estatus socioeconómico (ESCS).

$$\text{HOMEPOSS}_{ijk} = \gamma_{000} + \gamma_{100} \text{HEDRES}_{ijk} + \gamma_{200} \text{BAHEDRES}_{ijk} + \gamma_{300} \text{WEALTH}_{ijk} \\ + \gamma_{400} \text{BAWEALTH}_{ijk} + \gamma_{500} \text{CULTPOSS}_{ijk} + \gamma_{600} \text{BACULTPO}_{ijk} \\ + r_{0jk} + u_{00k} + e_{ijk}$$

Tabla 3.17 Resultados del modelo multinivel del índice HOMEPOS

Efecto fijo		Coefficiente	Error estándar	t	Grados de libertad	p-valor
Interceptos						
Valor base de HOMEPOSS	γ_{000}	0.2642	0.0136	19.44	6	0.000
Pendientes						
HEDRES	γ_{100}	0.2346	0.0113	20.71	30343	0.000
BAHEDRES	γ_{200}	-0.1442	0.0363	-3.97	30343	0.000
WEALTH	γ_{300}	0.7667	0.0145	52.87	30343	0.000
BAWEALTH	γ_{400}	0.4194	0.1318	3.18	30343	0.002
CULTPOSS	γ_{500}	0.3085	0.0066	46.73	30343	0.000
BACULTPO	γ_{600}	0.1117	0.0700	1.60	30343	0.110

Efecto aleatorio			Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
Nivel	Unidad							
1	Estudiante	e_{ijk}	0.2075	0.0430				93.8
2	Escuela	r_{0jk}	0.0517	0.0027	1081	1857.47	0.000	5.8
3	Región	u_{00k}	0.0136	0.0002	6	63.83	0.000	0.4

Desajuste (Deviance)	-7969.19
----------------------	----------

Se confirma que los coeficientes de todas las variables funcionan apropiadamente con el modelo multinivel en forma significativa. Por su parte, el desajuste es -7969, cuyo signo implica una imposibilidad del programa para estandarizar el resultado¹⁰, debido a la calidad de los datos disponibles, pero este resultado es aceptable porque indica un buen ajuste del modelo.

¹⁰ Puede consultarse la sección: "Log likelihood values and negative deviances", de la guía de ayuda del software HLM.

Por la definición de las variables, es evidente la nula participación del nivel 3 en la varianza de HOMEPOSS y resulta muy conveniente que el 95% de la varianza de la variable de interés en esta etapa del análisis se ubique en el nivel 1 para conservarla como variable explicativa en el modelo estructural del desempeño en Ciencias que se construirá más adelante.

De las tres variables utilizadas para modelar HOMEPOS, la de mayor influencia es WEALTH, que produce una diferencia de hasta 4.49 puntos entre casos extremos. Vale la pena recordar que el rango de valores de HOMEPOS es de 9.77 y su desviación estándar es 1.16, por lo que WEALTH cubre más de la mitad de lo que puede ocurrir con la información acerca de los bienes en el hogar. La posesión de bienes culturales es significativa, aunque cae en la categoría de mediana para esta variable.

Tabla 3.18 Valores de d^* para las variables del modelo HOMEPOS

Variable explicativa	Coeficiente	Categoría para resultado inferior	Categoría resultado superior	Diferencia γ Variable	d^*	Significancia y clasificación
WEALTH	0.7667	-3.96	1.9	4.49	19.37	Grande
HEDRES	0.2346	-4.64	1.08	1.34	5.78	Grande
CULTPOSS	0.3085	-1.44	1.08	0.78	3.35	Mediana

3.5.5 Ecuaciones del modelo multinivel

Al producir el modelo de *Eficiencia* con las variables del nivel 1 se incluye HOMEPOS como variable explicativa del estudiante, su hogar y su familia y se agregan las variables del nivel 2 identificadas en el componente factorial.

El modelo incluye estas variables:

$$\begin{aligned}
 PVSCIE_{ijk} = & \gamma_{000} + \gamma_{010} CLSIZE_{jk} + \gamma_{020} URBANORURAL_{jk} + \gamma_{030} IRATCOMP_{jk} \\
 & + \gamma_{040} STRATIO_{jk} + \gamma_{050} IRRESPCU_{jk} + \gamma_{100} HOMEPOS_{ijk} \\
 & + r_{0jk} + u_{00k} + e_{ijk}
 \end{aligned}$$

Las variables explicativas empleadas en este modelo intervienen de manera estadísticamente significativa, exceptuando STRATIO.

Tabla 3.19 Resultados del modelo multinivel de eficiencia

Efecto fijo		Coefficiente	Error estándar	t	Grados de libertad	p-valor
Interceptos						
Puntuación base en Ciencias	γ_{000}	356.887	8.09	44.11	6	0.000
CLSIZE	γ_{010}	1.075	0.21	5.19	83	0.000
URBANORURAL	γ_{020}	26.378	7.17	3.68	356	0.000
IRATCOMP	γ_{030}	107.287	15.85	6.77	1025	0.000
STRATIO	γ_{040}	0.508	0.29	1.74	127	0.084
IRRESPCU	γ_{050}	24.334	4.06	6.00	1082	0.000
Pendientes						
HOMEPOS	γ_{100}	4.661	0.92	5.05	10	0.000

Efecto aleatorio			Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
Nivel	Unidad							
1	Estudiante	e_{ijk}	58.35	3404.35				60.8
2	Escuela	r_{ojk}	45.39	2059.93	1076	16798.18	0.000	36.8
3	Región	u_{00k}	11.74	137.76	6	72.62	0.000	2.5

Desajuste (Promedio)	336159.95	Mejora respecto del modelo de referencia	0.15%	
Diferencia respecto del modelo nulo	505.18	Diferencia en grados de libertad	6	Diferencia significativa

No.	Nivel		Componente de varianza en el modelo nulo	Componente de varianza en este modelo	Diferencia de varianza %
1	Estudiante	e_{ijk}	3416.98	3404.35	0.4
2	Escuela	r_{ojk}	3063.17	2059.93	32.8
3	Región	u_{00k}	441.09	137.76	68.8

En este modelo se muestra la poca significancia del coeficiente de la variable STRATIO, con un modelo que mejora al incondicional en más de 0.15%, con una diferencia significativa.

3.5.6 Presentación e interpretación de resultados

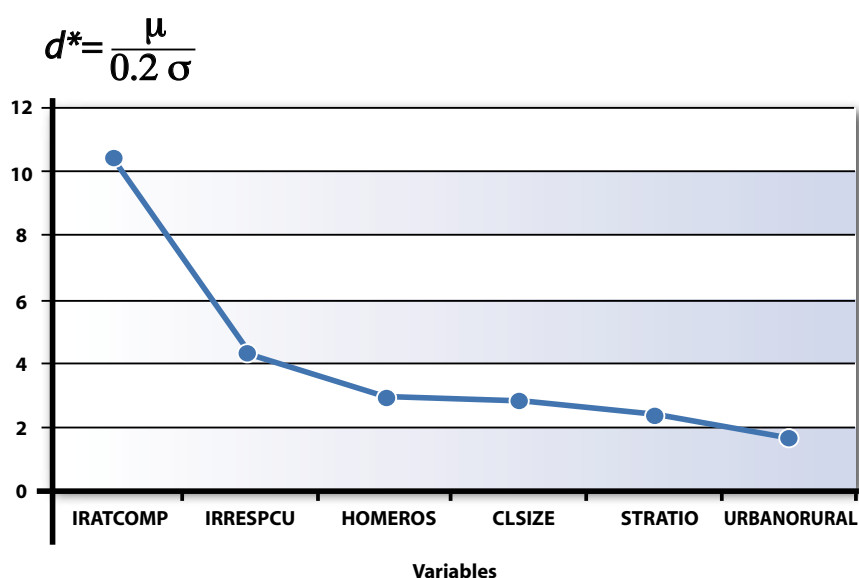
Se procede como en casos anteriores a identificar la diferencia de puntuación entre los casos extremos (Tabla 3.20):

Tabla 3.20 Valores de d^* para las variables del modelo de eficiencia

Variable explicativa	Coefficiente	Categoría para resultado inferior	Categoría resultado superior	Diferencia γ Variable	d^*	Significancia y clasificación
CLSIZE	1.075	13	53	43.00	2.84	Mediana
URBANORURAL	26.378	Rural = 0	Urbana = 1	26.38	1.74	Pequeña
IRATCOMP	107.287	0	1.47	157.71	10.42	Grande
STRATIO	0.508	0.27	73.43	37.17	2.45	Pequeña
IRRESPCU	24.334	-1.4	1.27	64.97	4.29	Grande
HOMEPOS	4.661	-6.54	3.23	45.54	3.01	Mediana
Total teórico (valor absoluto)				374.76	24.75	Grande
Equivalente en desviaciones estándar de Ciencias				4.95		
Equivalente respecto de la escala de puntuaciones en Ciencias %				62.46		

La gráfica presenta los valores de d^* una vez ordenados los valores en forma descendente. Todas las variables (de niveles 1 y 2) producen diferencias significativas ($d^* > 1$) en el desempeño de las categorías o valores extremos. Los casos con diferencias pequeñas ocurren en URBANORURAL y STRATIO.

Figura 3.14 Valores de d^* para las variables del modelo de eficiencia



3.5.7 Conclusiones e implicaciones de los resultados

Todas las variables consideradas en la dimensión de eficiencia inciden en la puntuación de los estudiantes en Ciencias de la prueba PISA, a tal grado que su combinación implica una diferencia muy grande entre valores extremos, abarcando 62% de la escala total de puntuaciones.

A nivel del estudiante, la disponibilidad de bienes culturales y otros satisfactores en el hogar inciden positivamente en más de 45 puntos de la escala entre los alumnos más extremos. Ya se sabe que dentro de ella, la variable explicativa WEALTH es la que tiene mayor participación en el modelado de HOMEPOS, se trata de una variable sobre la cual es difícil establecer esquemas de la escuela, estatales o regionales que pudieran modificarla.

Es posible incidir de manera relativamente sencilla en las otras dos variables: HEDRES y CULTPOSS, porque la familia, y el estudiante mismo, podrían allegarse de mayor número de bienes que mejoraran su nivel cultural; por ejemplo, se podría disponer de una mayor dotación de libros de lectura general y de consulta, además de materiales educativos (desde enciclopedias hasta recursos informáticos). No se trata de incrementar artificialmente el índice administrando colecciones de libros y de discos al estudiante, sino propiciar su uso y manejo, ya que la inversión de este tipo de materiales no es necesariamente más cara que la adquisición de otro tipo de bienes de consumo cotidiano.

A nivel de escuela, la proporción de computadoras para fines educativos parece ser la variable de mayor importancia en el desempeño de los estudiantes, con una diferencia $d^*=10.42$. El coeficiente obtenido de esta variable es alto, produciendo diferencias de 157 puntos en la escala PISA entre las categorías extremas.

Se sabe que una realidad mexicana es la sobrepoblación en las aulas y se argumenta que esto desfavorece el aprendizaje, pero los resultados no indican que en México el tamaño de la clase sea un problema. El tamaño promedio de la clase incide de manera mediana en el valor de $d^*=2.84$. La media actual reportada en México para el proyecto PISA es de 40.77 estudiantes y la pendiente de esta variable explicativa es positiva (1.075), lo cual indica que aparentemente no hay una desventaja en el hecho de contar con muchos alumnos en el salón de clase.

Si se acepta que la media de estudiantes es del orden de 40 por clase, ¿qué opciones tiene el docente para promover el aprendizaje? Hay varias soluciones, pero en este momento es posible pensar que cuando el grupo de clases es numeroso, el método de enseñanza-aprendizaje es de tipo expositivo principalmente, con lo que se limita la forma de trabajar en el aula y se restringe el desarrollo de los procesos reflexivos o constructivos y de las competencias exploradas en la prueba PISA. Esta hipótesis podría explorarse a futuro al combinar esta variable con otras que exploren el tipo de enseñanza-aprendizaje promovido en clase.

Sugerencias

Esta variable conlleva varias posibles líneas de trabajo e investigación.

1. Comparar el consumo que tiene el estudiante de diversos productos y servicios (cigarrillos, revistas de poco impacto cultural, bebidas gaseosas, comida *chatarra*, renta de equipo de cómputo en un café Internet o uso de la computadora en caso de tenerla disponible en casa, adquisición de productos pirata, renta de películas en un club de video) contra otro tipo de productos de mejor contenido alimenticio y bienes culturales. La comparación no es trivial, porque un estudiante de escasos recursos no tiene, posiblemente, el dinero suficiente para satisfacer sus mínimas necesidades de subsistencia y debido a ello difícilmente puede considerar la adquisición de bienes culturales para sí mismo o para el hogar; pero en otros estratos es muy probable que el estudiante tenga la opción de elegir entre una gama de productos y, simultáneamente, disponer del dinero suficiente para adquirir y consumir productos que le aporten a su desempeño académico. Para esta comparación se deben elaborar instrumentos que exploren formas de consumo de los estudiantes con los cuales establecer comparativos de preferencias y la manera de integrar el índice HOMEPOS.
2. Plantear campañas de promoción cultural, distribuir libros educativos de bajo costo y verificar al mismo tiempo su empleo, para construir un índice de eficiencia en el uso de los recursos para la mejora en el desempeño de los estudiantes.
3. Tomando en cuenta que IRATCOMP es una variable sobre la cual se puede incidir de manera relativamente fácil, se deberá revisar el uso de los recursos informáticos para mejorar el aprendizaje de los estudiantes. Por ejemplo, si se cuenta con una computadora por cada cinco estudiantes ($\text{IRATCOMP}=0.2$) se tiene un efecto de 21 puntos de la escala PISA (107×0.2), actualmente la media de IRATCOMP es de 0.08 (aproximadamente una computadora por cada 12 estudiantes), con un efecto de 8.5 puntos en la escala PISA. Un análisis beneficio-coste por institución permitiría establecer la mejor política de disponibilidad y uso de equipo informático, para obtener el máximo de desempeño en los estudiantes.

3.6 Modelos para la dimensión de impacto

3.6.1 Pregunta de investigación

Las preguntas relativas al impacto en el Modelo de Calidad son estas:

- ¿Un mayor desempeño en Ciencias puede asociarse con un considerable empleo y conciencia en la vida cotidiana del estudiante?
- Si los alumnos cuentan con actitudes para la ciencia y dominan algunos rasgos no cognitivos que las favorezcan, ¿se mejora el desempeño en Ciencias?

3.6.2 Hipótesis

El desempeño de los estudiantes en la prueba PISA puede asociarse significativamente con diversas manifestaciones y actitudes no cognitivas.

3.6.3 Variables explicativas

Las variables de los niveles 1 y 2 que intervienen en los componentes factoriales son los siguientes:

Tabla 3.21 Variables del modelo de impacto

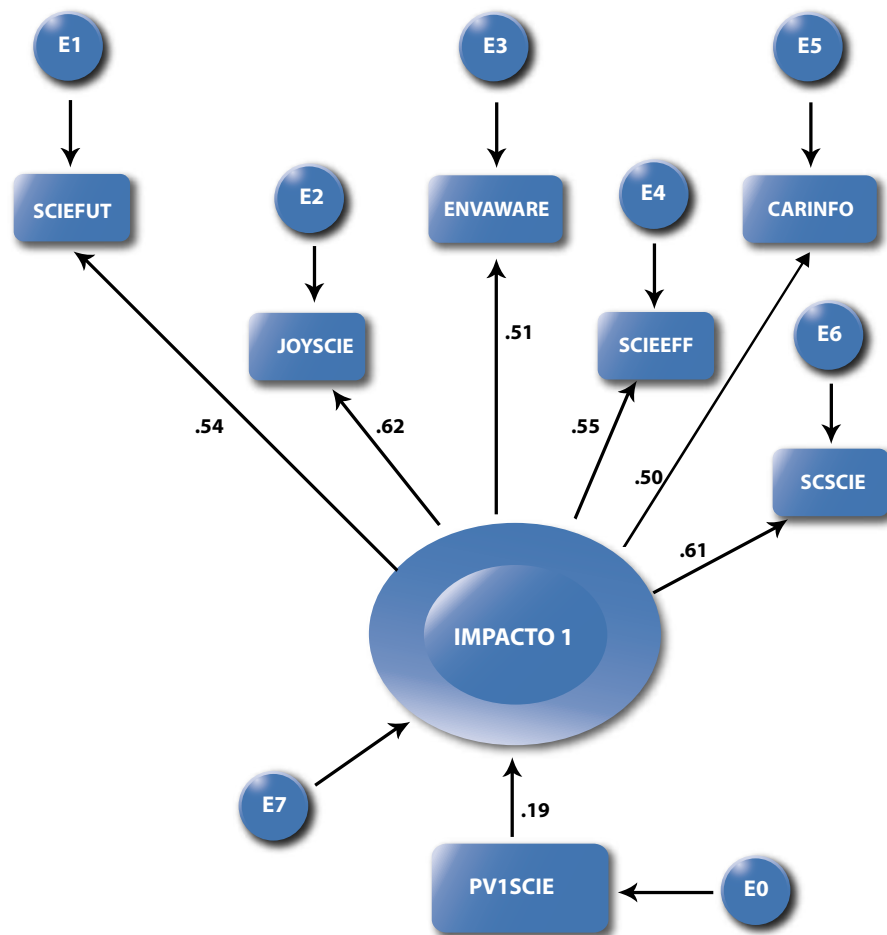
Nivel	Variables	Descripción
Variables de nivel 1	ENVAWARE	Conocimiento de aspectos medio ambientales
	SCIEEFF	Autoeficacia en relación con las Ciencias
	CARINFO	Información sobre carreras relacionadas con las Ciencias
	JOYSCIE	Disfrute al aprender temas relacionadas con las ciencias
	SCIEFUT	Motivación orientada hacia las ciencias
	SCSCIE	Autoconcepto en ciencias
	RESPDEV	Conciencia del estudiante para el desarrollo sustentable
	GENSCIE	Valor que tiene la ciencia para el estudiante
	CARPREP	Formación que proporciona la escuela para las carreras relacionadas con las Ciencias
	PERSCIE	Valor personal del estudiante hacia las ciencias

3.6.4 Resultados del análisis factorial confirmatorio

Todas las variables se ubican en el nivel 1 y forman dos componentes en el análisis factorial exploratorio. Para los modelos estructurales se asociaron con variables latentes mediadoras denominadas IMPACTO 1 e IMPACTO 2.

Al asociar las variables de nivel 1 dentro de un primer tipo de impacto (ajuste aceptable AGFI=0.841), se tienen coeficientes de contribución muy similares entre sí (de 0.5 a 0.62) pero, al mismo tiempo, se tiene un efecto pequeño de 0.19 en el coeficiente de desempeño en Ciencias (figura 3.15). La variable latente de impacto incluye aspectos no cognitivos donde el estudiante establece su gusto por las Ciencias, su percepción de aplicaciones en el futuro y su forma de valorar la importancia de las ciencias en el medio ambiente, junto con aspectos personales sobre su nivel de información y de auto eficacia en el uso de las Ciencias.

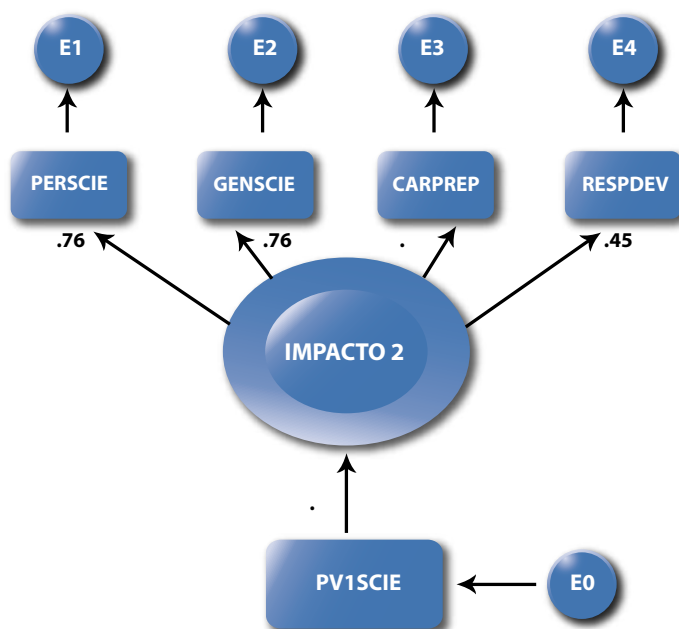
Figura 3.15 Modelo estructural de la dimensión de impacto I



AGFI=0.841

Como resultado del análisis factorial se aisló un segundo tipo de impacto (ajuste aceptable 0.872), formado por las variables PERSCIE, GENSCIE, CARPREP y RESPDEV que, por parejas tienen contribuciones similares (0.76 en una pareja y 0.44 en la otra), aunque el efecto es menor en el desempeño en Ciencias que el otro conjunto de variables de impacto.

Figura 3.16 Modelo estructural de la dimensión de impacto 2



AGFI=0.872

3.6.5 Ecuaciones del modelo multinivel

Se prepararon dos modelos para la dimensión de impacto, siguiendo los modelos estructurales descritos previamente con base en las variables del nivel I.

$$\begin{aligned} PVSCIE_{ijk} = & \gamma_{000} + \gamma_{100} ENVAWARE_{ijk} + \gamma_{200} SCIEEFF_{ijk} + \gamma_{300} CARINFO_{ijk} \\ & + \gamma_{400} JOYSCIE_{ijk} + \gamma_{500} SCIEFUT_{ijk} + \gamma_{600} SCSCIE_{ijk} \\ & + r_{0jk} + u_{00k} + e_{ijk} \end{aligned}$$

$$\begin{aligned} PVSCIE_{ijk} = & \gamma_{000} + \gamma_{100} RESPDEV_{ijk} + \gamma_{200} GENSCIE_{ijk} + \gamma_{300} CARPREP_{ijk} \\ & + \gamma_{400} PERSCIE_{ijk} + r_{0jk} + u_{00k} + e_{ijk} \end{aligned}$$

Las variables explicativas empleadas en este modelo intervienen de manera estadísticamente significativa, excepto SCIEFUT y SCSCIE en el primer modelo de impacto.

Tabla 3.22 Resultados del primer modelo de impacto

Efecto fijo		Coeficiente	Error estándar	t	Grados de libertad	p-valor
Interceptos						
Puntuación base en Ciencias	γ_{000}	417.046	8.18	50.99	6	0.000
Pendientes						
ENVAWARE	γ_{100}	17.203	1.26	13.62	6	0.000
SCIEEFF	γ_{200}	6.025	1.09	5.52	41	0.000
CARINFO	γ_{300}	-6.935	0.62	-11.13	7	0.000
JOYSCIE	γ_{400}	8.990	2.01	4.48	108	0.000
SCIEFUT	γ_{500}	-2.581	2.23	-1.16	1871	0.247
SCSCIE	γ_{600}	0.944	1.93	0.49	27	0.629

Efecto aleatorio			Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
Nivel	Unidad							
1	Estudiante	e_{ijk}	55.28	3056.35				52.2
2	Escuela	r_{0jk}	49.61	2461.06	1081	23423.78	0	42.0
3	Región	u_{00k}	18.37	337.47	6	203.54	0	5.8

No.	Nivel		Componente de varianza en el modelo nulo	Componente de varianza en este modelo	Diferencia de varianza %
1	Estudiante	e_{ijk}	3416.98	3056.35	10.6
2	Escuela	r_{0jk}	3063.17	2461.06	19.7
3	Región	u_{00k}	441.09	337.47	23.5

Desajuste (Promedio)	333171.91	Mejora respecto del modelo de referencia	1.04%	
Diferencia respecto del modelo nulo	3493.22	Diferencia en grados de libertad	6	Diferencia significativa

La proporción de varianza es comparable entre los niveles 1 y 2, con muy poca participación del nivel 3 (menos de 6% en ambos modelos). La mejora del modelo 1 respecto del caso incondicional es mayor que la que se obtiene con el modelo 2 (1.04% y 0.62% respectivamente). El primer modelo de la dimensión de impacto es un poco más explicativo de la varianza en los tres niveles (10.6, 19.7 y 23.5% respectivamente) que el segundo modelo (6.5, 11.8 y 17.7%).

Tabla 3.23 Resultados del segundo modelo de impacto

Efecto fijo		Coeficiente	Error estándar	t	Grados de libertad	p-valor
Interceptos						
Puntuación base en Ciencias	γ_{000}	414.469	9.51	43.59	6	0.000
Pendientes						
RESPDEV	γ_{100}	13.300	1.99	6.68	5432	0.000
GENSCIE	γ_{200}	9.933	1.26	7.88	22	0.000
CARPREP	γ_{300}	-7.268	1.41	-5.15	66	0.000
PERSCIE	γ_{400}	-3.659	1.00	-3.65	76	0.001

Efecto aleatorio	Nivel	Unidad		Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
1	Estudiante	e_{ijk}		56.53	3195.11				51.1
2	Escuela	r_{0jk}		51.96	2700.23	1081	23118.10	0.000	43.1
3	Región	u_{00k}		19.05	362.87	6	184.38	0.000	5.8

No.	Nivel		Componente de varianza en el modelo nulo	Componente de varianza en este modelo	Diferencia de varianza %
1	Estudiante	e_{ijk}	3416.98	3195.11	6.5
2	Escuela	r_{0jk}	3063.17	2700.23	11.8
3	Región	u_{00k}	441.09	362.87	17.7

Desajuste (Promedio)	334567.81	Mejora respecto del modelo de referencia	0.62 %	
Diferencia respecto del modelo nulo	2097.32	Diferencia en grados de libertad	4	Diferencia significativa

3.6.6 Presentación e interpretación de resultados

Cada modelo funciona independientemente, es decir, no son aditivos aquéllos de impacto 1 y 2, por lo que el interés de cada una de las variables en el desempeño se tiene que estudiar para cada caso. El primer modelo contiene el mayor efecto, de 110.96 puntos en la variable ENVAAWARE (tabla 3.24); es razonable pensar que los estudiantes perciben a las Ciencias como herramienta para resolver problemas ambientales y esto ayuda a tener un mejor desempeño en la prueba PISA. El mismo modelo revela que para tener un mejor desempeño, no cuenta de manera significativa que el alumno tenga una motivación para continuar sus estudios o trabajar en áreas científicas, ni tampoco que tenga una buena apreciación personal de su potencial en tales asignaturas (SCIEFUT y SCSCIE).

Tabla 3.24 Valores de d^* para las variables que integran el primer modelo de impacto

Variable explicativa	Coeficiente	Categoría para resultado inferior	Categoría resultado superior	Diferencia γ Variable	d^*	Significancia y clasificación
ENVAWARE	17.203	-3.44	3.01	110.96	7.33	Grande
SCIEEFF	6.025	-3.77	3.22	42.11	2.78	Mediana
CARINFO	-6.935	-2.44	2.53	-34.47	-2.28	Pequeña
JOYSCIE	8.99	-2.15	2.06	37.85	2.50	Pequeña
SCIEFUT	-2.581	-1.42	2.27	-9.52	-0.63	Muy pequeña
SCSCIE	0.944	-2.36	2.24	4.34	0.29	Muy pequeña
Total teórico (valor absoluto)				239.26	15.80	Grande
Equivalente en desviaciones estándar de Ciencias				3.16		
Equivalente respecto de la escala de puntuaciones en Ciencias %				39.88		

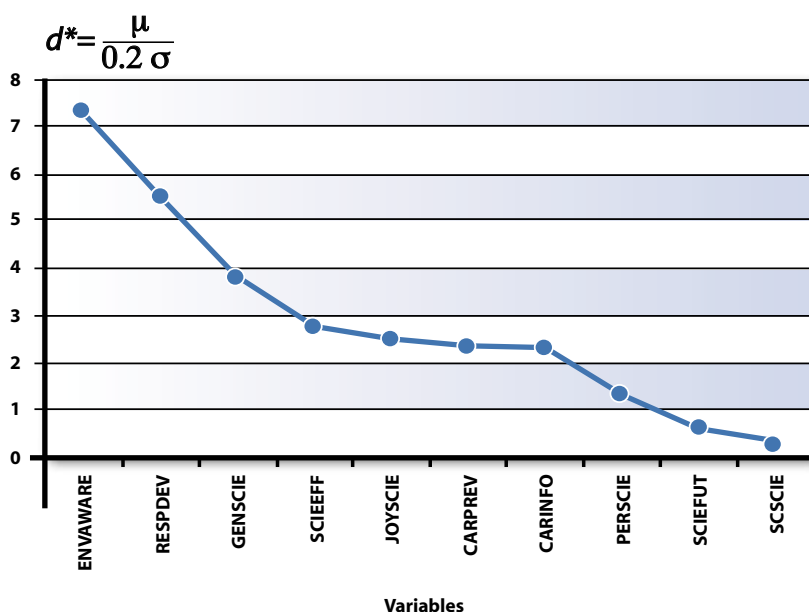
Resulta inesperado que el hecho de disponer de información sobre carreras científicas tenga un coeficiente de impacto negativo en el desempeño en Ciencias y llegue a producir diferencias significativas entre valores extremos de la variable.

El mayor efecto en el segundo modelo ocurre con RESPDEV, donde se puede tener una diferencia de 83.79 puntos entre valores extremos de la escala PISA (tabla 3.25). Esta es una variable actitudinal hacia cuestiones ambientales que favorecen el desarrollo sustentable; la magnitud del coeficiente indica que los estudiantes con mayor enfoque ambientalista tienen un mejor desempeño en Ciencias.

Tabla 3.25 Valores de d^* para las variables que integran el segundo modelo de impacto

Variable explicativa	Coeficiente	Categoría para resultado inferior	Categoría resultado superior	Diferencia γ Variable	d^*	Significancia y clasificación
RESPDEV	13.3	-4	2.3	83.79	5.53	Grande
GENSCIE	9.933	-3.66	2.19	58.11	3.84	Mediana
CARPREP	-7.268	-2.92	1.96	-35.47	-2.34	Pequeña
PERSCIE	-3.659	-3.08	2.53	-20.53	-1.36	Pequeña
Total teórico (valor absoluto)				197.89	13.07	Grande
Equivalente en desviaciones estándar de Ciencias				2.61		
Equivalente respecto de la escala de puntuaciones en Ciencias %				32.98		

Figura 3.17 Valores de d^* para las variables de la dimensión de impacto



3.6.7 Conclusiones e implicaciones de los resultados

El análisis multinivel indica que el impacto que se asocia con el desempeño en Ciencias es importante en la mayoría de las variables consideradas. Con excepción de las variables SCSCIE y SCIEFUT.

Las variables ENVAWARE y RESPDEV son las más importantes para producir diferencias significativas dentro de la categoría de *grande* para d^* . Las distribuciones de los valores de ambas variables tienen desviaciones estándar de 0.9, pudiéndose comprobar que un cambio de una desviación estándar en la atención al medio ambiente y al desarrollo sustentable pueden producir respectivamente 17 y 13 puntos en la escala de PISA. La diferencia observada en el impacto que producen ambas variables puede deberse a que RESPDEV se construye con preguntas que exploran aspectos actitudinales, mientras que ENVAWARE se integra con preguntas sobre aspectos cognitivos.

Los efectos que producen estas dos variables son menos importantes que los producidos por otras variables como el ESCS visto previamente en el modelo de equidad, pero corresponden a valores de d^* de 7.33 y 5.33, que entran en la categoría de *grande* y, por tanto, son significativos. Se trata entonces de variables que con un pequeño esfuerzo de promoción por parte de la institución o de los docentes, pueden aportar un valor agregado de interés en el desempeño de los estudiantes en Ciencias.

Hay cuatro variables donde el impacto interviene en forma inversa, se trata de CARINFO, SCIEFUT, CARPREV y PERSCIE; el origen del coeficiente negativo puede tener una causa similar a la planteada con las variables de pertinencia, es decir, que los estudiantes no comprenden lo que se pregunta o que intentan dar respuestas *socialmente deseables*. En caso de que no se puedan verificar ambas conjeturas, se cae en

la problemática de que los estudiantes muestran actitudes adversas hacia las Ciencias y que la información que reciben por diversos medios, incluyendo la escuela, tiene un efecto contraproducente e inhibidor del interés por las asignaturas científicas.

Sugerencia

Para promover los temas científicos y sus aplicaciones entre los estudiantes se recomiendan actividades que motiven y desarrollen un interés positivo por esta área del conocimiento. Asimismo, se sugiere verificar las conjeturas planteadas sobre las variables de coeficientes negativos.

1. Si se parte de que el dominio de competencias en Ciencias se refleja en un impacto considerable en la manera de comprender los fenómenos de la naturaleza y de la vida cotidiana, materializándose en formas de actuar ante problemas reales, entonces resulta muy conveniente promover las aplicaciones de las Ciencias tanto en los aspectos cognitivos como en los actitudinales. Se deben aprovechar las asignaturas científicas para despertar una conciencia ecológica, de cuidado del medio ambiente y proactiva hacia el desarrollo sustentable.
2. Definir mecanismos de mediación pedagógica u otro modelo de intervención en el aula y a nivel escolar y estatal, para promover el aprendizaje y la utilización de las Ciencias; los estudiantes deben reconocer las oportunidades profesionales que representa el campo científico y, para ello, deben contar con información sobre carreras, aunado al desarrollo de valores y la mejora de las actitudes hacia el estudio de áreas científicas.
3. Discriminar si los coeficientes negativos que se presentan en las variables CARIN-FO, SCIEFUT, CARPREP y PERSCIE ocurren por deficiencias en la comprensión de las preguntas o por el propósito de proporcionar respuestas *socialmente deseadas*. En cualquiera de los dos casos se deberán proponer acciones correctivas para futuras aplicaciones de cuestionario de contexto. Si el signo de los coeficientes es producto de otras razones, se deberá realizar un trabajo de investigación más detallado que permita dilucidar el origen de esta tendencia anómala.

3.7 Modelo de la dimensión de eficacia

3.7.1 Pregunta de investigación

¿Puede afirmarse que las variables socioeconómicas de tercer nivel son eficaces para incidir en el desempeño de los estudiantes en la prueba de Ciencias de PISA?

3.7.2 Hipótesis

Hay variables regionales de corte socioeconómico (nivel 3) que tienen eficacia para incidir significativamente en el desempeño de los estudiantes en la prueba de Ciencias del proyecto PISA.

3.7.3 Variables explicativas

La base de datos del nivel 3 fue construida con variables socioeconómicas, demográficas y académicas, estandarizadas con media 0 y desviación estándar 1. El análisis factorial produce un componente principal que se identificó relacionando la dimensión de eficacia del Modelo de Calidad Educativa del INEE. Esta dimensión presenta algunas diferencias con los otros modelos presentados, porque varias de las variables están altamente correlacionadas desde su definición, como es el caso del PIB y el IDH que tienen una relación funcional entre ellas. Como se verá más adelante, los modelos de esta dimensión incluyen pocas variables de nivel 3, porque los análisis se tuvieron que fraccionar para obtener la convergencia en el *software* HLM.

Tabla 3.26 Variables del modelo de eficacia

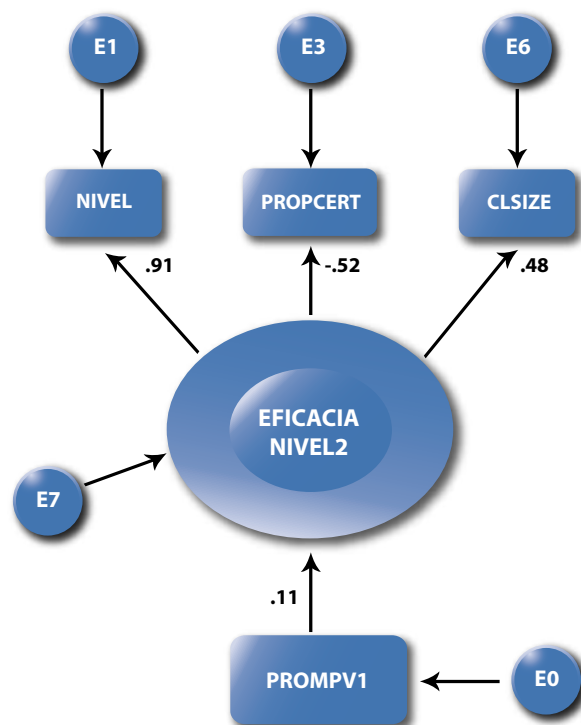
Nivel	Variables	Descripción
Variables de nivel 2	NIVEL	Nivel de estudios que imparte la escuela 0= Secundaria 1= Educación media superior
	PROPCERT	Proporción de los profesores con licenciatura.
	CLSIZE	Promedio del tamaño de clase
Variables de nivel 3	PIBPERCA	Producto interno bruto <i>per cápita</i> , en dólares PPA, en 2004
	IDH	Índice de Desarrollo Humano
	GRSOBRE	Cociente del grado de escolaridad entre el analfabetismo
	MARGINAL	Índice de marginación en 2005
	COBERTUR	Tasa neta de cobertura en 2005
	COMPUPUB	Porcentaje de escuelas públicas sin computadora
	COMPUPRI	Porcentaje de escuelas privadas sin computadora

3.7.4 Resultados del análisis factorial confirmatorio

El nivel 3 agrupa la mayoría de las variables en un gran componente factorial, debido especialmente a la alta correlación entre los datos socioeconómicos. En el nivel 2 se tienen tres variables de la escuela que también pueden agruparse en un componente factorial que se identificó como parte de la dimensión de eficacia: NIVEL, PROPCERT y CLSIZE.

El modelo factorial confirmatorio realizado con ecuaciones estructurales (con ajuste muy bueno AGFI=0.998) incluye una variable mediadora latente denominada EFICACIANIVEL2, que se asocia con un coeficiente de 0.11 con el desempeño en Ciencias. En las otras variables, el cálculo de las ecuaciones estructurales indica que el coeficiente estandarizado más alto se tiene en la variable NIVEL, en tanto que las otras variables tienen coeficientes cercanos en valor absoluto (0.48, 0.52), pero la proporción de docentes certificados interviene en forma negativa, lo cual es un resultado inesperado.

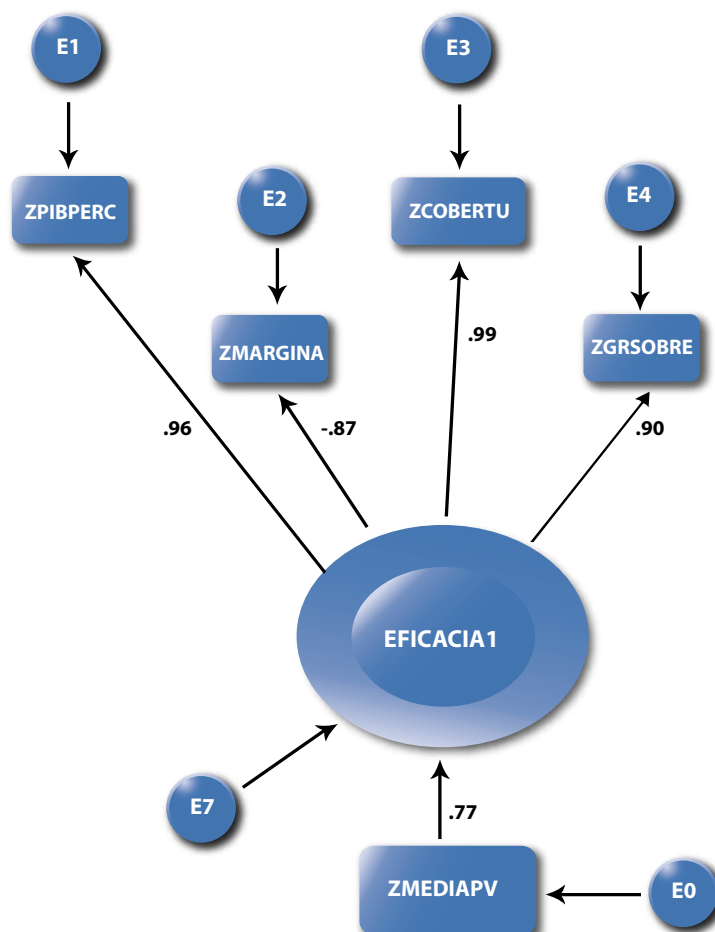
Figura 3.18 Modelo estructural de la dimensión de eficacia nivel 2



AGFI=0.998

Para el nivel 3, el modelo propuesto para la variable latente EFICACIA incluye PIB *per cápita*, MARGINAL y COBERTURA, además de la variable combinada del grado escolar sobre analfabetismo. Como era de esperarse, la marginalidad participa de manera negativa, indicando que a mayor marginalidad se tiene un menor desempeño en la prueba PISA. Todas las demás variables intervienen de manera positiva con coeficientes de influencia muy altos (0.90 a 0.99). Por la cantidad tan reducida de registros de este nivel no se obtiene un buen resultado en el parámetro AGFI y se reporta el CFI=0.881 que es un ajuste todavía aceptable.

Figura 3.19 Modelo estructural de la dimensión de eficacia nivel 3



CFI=0.881

Las variables que intervienen en estos análisis están altamente correlacionadas, como es el caso de IDH y PIB *per cápita*. Estas altas correlaciones inciden en la colinealidad de las variables y en la no convergencia de los modelos multinivel. Por esta razón y porque se tienen muy pocos registros, no es posible establecer modelos multinivel con más de dos variables explicativas del nivel 3, para probar la dimensión de eficacia. En consecuencia, los modelos obtenidos con variables del nivel 3 deben considerarse como una aproximación a la dimensión de eficacia.

3.7.5 Ecuaciones del modelo multinivel

Se dividen las variables de nivel 3 en dos grupos, para tener soluciones convergentes, por ejemplo:

$$PVSCIE_{ijk} = \gamma_{000} + \gamma_{001} ZPIBPERC_k + \gamma_{002} ZMARGINA_k + r_{0jk} + u_{00k} + e_{ijk}$$

$$PVSCIE_{ijk} = \gamma_{000} + \gamma_{001} ZGRSOBRE_k + \gamma_{002} ZCOBERTU_k + r_{0jk} + u_{00k} + e_{ijk}$$

(Solo se presentan los resultados del primero de estos modelos).

Un modelo factible que combina variables de los niveles 2 y 3 es:

$$PVSCIE_{ijk} = \gamma_{000} + \gamma_{001} ZPIBPERC_k + \gamma_{002} ZMARGINA_k + \gamma_{010} CLSIZE_{jk} + \gamma_{020} PROPCERT_{jk} + \gamma_{030} NIVEL_{jk} + r_{0jk} + u_{00k} + e_{ijk}$$

Las dos variables explicativas de nivel 3 del primer modelo intervienen de manera estadísticamente significativa (tabla 3.27). Es notable que esta combinación explica prácticamente la totalidad de la varianza del nivel 3 (componente de varianza de 0.66 y 99.9% de diferencia respecto del modelo incondicional), aunque no aportan ninguna mejora en la explicación de la varianza de los niveles 1 y 2.

Tabla 3.27 Resultados del modelo de la dimensión de eficacia con variables de nivel 3

Efecto fijo		Coefficiente	Error estándar	t	Grados de libertad	p-valor
Interceptos						
Puntuación base global en Ciencias	γ_{000}	404.236	1.50	270.13	4	0.000
ZPIBPERC	γ_{001}	11.711	1.70	6.88	4	0.000
ZMARGINA	γ_{002}	-10.290	1.94	-5.30	4	0.002

Efecto aleatorio			Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
Nivel	Unidad							
1	Estudiante	e_{ijk}	58.45	3416.91				52.8
2	Escuela	r_{0jk}	55.24	3051.88	1081	24934.72	0.000	47.2
3	Región	u_{00k}	0.81	0.66	4	9.86	0.042	0.0

Desajuste (Promedio)	336638.53	Mejora respecto del modelo de referencia	0.01%	
Diferencia respecto del modelo nulo	26.60	Diferencia en grados de libertad	2	Diferencia significativa

No.	Nivel		Componente de varianza en el modelo nulo	Componente de varianza en este modelo	Diferencia de varianza %
1	Estudiante	e_{ijk}	3416.98	3416.91	0.0
2	Escuela	r_{0jk}	3063.17	3051.88	0.4
3	Región	u_{00k}	441.09	0.66	99.9

Este modelo se presentó sólo con variables del nivel 3 (ver tabla 3.28), para mostrar que todas son estadísticamente significativas, pero cuando se combinan variables de los niveles 2 y 3, la marginalidad deja de ser significativa ($p=0.092$) y la proporción de docentes certificados conduce a un coeficiente negativo, pero no estadísticamente significativo. El conjunto de variables produce una diferencia en la varianza del nivel 3 de 99.8% y mejora la explicación en el nivel 2 con una diferencia de 29.3%; en el nivel 1 no se tiene una mejora explicativa de la varianza (0.1%),

Tabla 3.28 Resultados del modelo de la dimensión de eficacia con variables de los niveles 2 y 3

Efecto fijo		Coeficiente	Error estándar	t	Grados de libertad	P-valor
Interceptos						
Puntuación base global en Ciencias	γ_{000}	376.396	9.38	40.13	4	0.000
ZPIPERC	γ_{001}	11.227	2.65	4.24	4	0.019
ZMARGINA	γ_{002}	-5.599	2.58	-2.17	4	0.092
CLSIZE	γ_{010}	0.389	0.18	2.21	117	0.029
PROPCERT	γ_{020}	-16.420	11.91	-1.38	1084	0.168
NIVEL	γ_{030}	43.857	13.65	3.21	1084	0.002

Efecto aleatorio			Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
Nivel	Unidad							
1	Estudiante	e_{ijk}	58.44	3414.95				61.2
2	Escuela	r_{ojk}	46.54	2165.52	1078	15573.70	0.000	38.8
3	Región	u_{ook}	1.00	1.00	4	15.16	0.005	0.0

No.	Nivel		Componente de varianza en el modelo nulo	Componente de varianza en este modelo	Diferencia de varianza %
1	Estudiante	e_{ijk}	3416.98	3414.95	0.1
2	Escuela	r_{ojk}	3063.17	2165.52	29.3
3	Región	u_{ook}	441.09	1.00	99.8

Desajuste (Promedio)	336283.09	336297.76	Mejora respecto del modelo de referencia	0.11%	
Diferencia respecto del modelo nulo	382.04	367.37	Diferencia en grados de libertad	5	Diferencia significativa

3.7.6 Interpretación de resultados

Hay varios puntos relevantes de los modelos construidos para la dimensión de eficacia. En primer lugar, es importante observar por qué no se pueden incluir muchas más variables del nivel 3 en el modelo; esto se debe a la menor participación de la varianza en dicho nivel y a que las variables incluidas en el modelo logran explicar prácticamente la totalidad de su componente de varianza, obteniéndose diferencias

de 99.9% respecto de la varianza que se reporta en el modelo incondicional o modelo nulo. El modelo ya no acepta más variables para poder explicar la varianza del último nivel de anidamiento. También debe complementarse esta idea, apuntando que las variables socioeconómicas correlacionan muy fuertemente y son explicativas de las fluctuaciones en este nivel, haciendo poco factible combinar más de dos variables con estas características en el mismo modelo.

En segundo lugar, se observa que las variables elegidas en los niveles 2 y 3 son independientes dentro de su nivel de anidamiento, por lo que no explican prácticamente nada de la varianza en el nivel 1.

Tabla 3.29 Valores de d^* para las variables del modelo de eficacia 1

Variable explicativa	Coeficiente	Categoría para resultado inferior	Categoría resultado superior	Diferencia y Variable	d^*	Significancia y clasificación
ZPIBPERC	11.711	-0.87	2.07	34.43	2.27	Pequeña
ZMARGINA	-10.29	-1.34	1.75	-31.80	-2.10	Pequeña
Total teórico (valor absoluto)				66.226	4.37	Grande
Equivalente en desviaciones estándar de Ciencias				0.8749		
Equivalente respecto de la escala de puntuaciones en Ciencias %				11.038		

La participación del PIB *per cápita*, estadísticamente significativa (tabla 3.30), ocurre en la dirección esperada (coeficiente positivo), porque un incremento en este índice debe ser eficaz para mejorar el desempeño en Ciencias de la prueba PISA. La marginalidad, como pudiera preverse, tiene un coeficiente negativo, proporcionando una significancia pequeña en d^* para los valores extremos de la escala. La combinación de variables puede producir diferencias absolutas entre extremos que caen en la categoría grande ($d^*=4.37$).

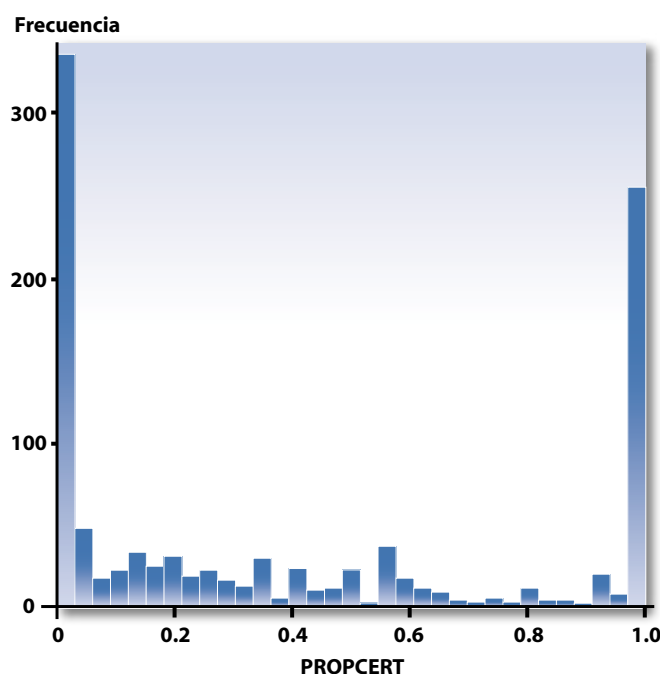
Tabla 3.30 Valores de d^* para las variables del modelo de eficacia 2

Variable explicativa	Coeficiente	Categoría para resultado inferior	Categoría resultado superior	Diferencia y Variable	d^*	Significancia y clasificación
ZPIBPERC	11.227	-0.87	2.07	33.01	2.18	Pequeña
ZMARGINA	-5.599	-1.34	1.75	-17.30	-1.14	Pequeña
CLSIZE	0.389	13	53	15.56	1.03	Pequeña
PROPCERT	-16.420	0	1	-16.42	-1.08	Pequeña
NIVEL	43.857	0	1	43.86	2.90	Mediana
Total teórico (valor absoluto)				126.15	8.33	Grande
Equivalente en desviaciones estándar de Ciencias				1.67		
Equivalente respecto de la escala de puntuaciones en Ciencias %				21.02		

El modelo multinivel con variables de los niveles 2 y 3 llega a un efecto combinado de 21% del rango total de la escala PISA, significativamente grande, aunque por separado cada variable tiene efectos significativamente pequeños, con excepción de NIVEL que es de efecto mediano.

El resultado inesperado en este modelo es que la proporción de docentes certificados tiene un coeficiente negativo. Este caso podría interpretarse erróneamente en el sentido de que los profesores certificados inciden en peores resultados en la prueba PISA, pero debe recordarse que no se dispone en este momento de una calificación de atributos profesionales o académicos de los docentes, sino exclusivamente la proporción de certificados. La razón del signo en el coeficiente se basa en la distribución de frecuencias de la proporción de docentes (variable continua de 0 a 1), que tiene una moda de mayor frecuencia en 0 que en 1 (337 escuelas con menos de 3% de docentes certificados contra 256 escuelas que tienen más de 97% de profesores certificados), con pocos casos en proporciones intermedias (figura 3.20). Sin duda, la mayor cantidad de escuelas con docentes sin certificación es la que produce el signo negativo en el coeficiente del modelo e indica que, con los datos disponibles, los resultados en PISA reportan el estado actual de la certificación docente, sobre la que habrá que trabajar, para evitar que se tenga este efecto negativo en el futuro.

Figura 3.20 Distribución de frecuencias de la proporción de docentes certificados en las escuelas



El tamaño de los grupos en las escuelas tiene un efecto pequeño, que confirma el resultado visto en el modelo de eficiencia, donde se comentó que el número de

alumnos en la clase no perjudica al desempeño de los estudiantes. Finalmente, el nivel educativo de los estudiantes es más eficaz sobre el desempeño, porque el coeficiente de 43.86 incide en una diferencia mediana.

Ya se ha comentado que hay una alta correlación de las variables socioeconómicas, lo cual hace complejo el trabajo de elección de variables que pueden participar en el modelo multinivel; para ello se utilizó el mismo *software* HLM6 que dispone de una opción para realizar un análisis exploratorio de las variables del nivel 3, que consiste básicamente en tomar a cada variable funcionando aisladamente dentro del modelo. Al término del análisis se obtiene el coeficiente estimado para la variable, el error estándar y el valor *t* (que debe ser superior a 2 para ser potencialmente significativo). Las variables de posible interés en el nivel 3 son las siguientes:

Tabla 3.3 I Variables potencialmente significativas en el nivel 3

	ZPIBPERC	ZIDH	ZGRSobre	ZMARGINA	ZCOMPUPU	ZCOMPUPR
Coeficiente	17.623	18.046	17.512	-17.713	-18.427	17.100
Error estándar	8.601	8.601	8.601	8.601	8.601	8.601
t	2.049	2.098	2.036	-2.059	-2.143	1.988

Se confirma la participación de algunas de las variables ya consideradas (PIB *per cápita*, IDH, grado de escolaridad sobre analfabetismo y marginalidad), además de la proporción de escuelas sin computadoras para uso académico en escuelas públicas y privadas.

3.7.7 Conclusiones e implicaciones de los resultados

Debe insistirse que los modelos de este estudio engloban variables de un componente factorial interpretado como parte de la dimensión de eficacia. Esta aproximación difiere de otros estudios (ver capítulo I) que utilizan criterios de jueceo para elegir a las variables, pudiendo encuadrar indistintamente variables de contexto, de eficiencia, de equidad y de eficacia.

Las variables socioeconómicas del nivel 3, identificadas dentro de la dimensión de eficacia, se asocian de manera interesante con el desempeño de los estudiantes en Ciencias de la prueba PISA. De este modo, un estudiante de educación media superior, que vive y estudia en un estado cuyo PIB *per cápita* es alto y su marginalidad es baja, puede tener 94 puntos de la escala PISA por arriba de otro estudiante que se encuentre en las condiciones más adversas (estudiante de secundaria, bajo PIB *per cápita* y alta marginalidad).

Las variables IDH, PIB *per cápita* y otras con las que están altamente correlacionadas, por tratarse de variables socioeconómicas estatales o regionales del nivel 3, pueden representar una fuente de información de interés y utilidad para políticas educativas nacionales, regionales o estatales.

Sugerencias

Las sugerencias para esta dimensión se enfocan a la escuela y al ámbito de acción de las autoridades estatales.

1. Se sugiere revertir la tendencia de la marginalidad, variable que participa en forma negativa y que puede afectar el desempeño de los estudiantes en Ciencias en más de 30 puntos. No es evidente definir un conjunto de acciones para reducir o eliminar la marginalidad, pero se puede actuar sobre su posible participación en los resultados, se trata de establecer acciones enfocadas a apoyar a los estudiantes de menores recursos, especialmente en los estados con mayor nivel de marginación.
2. Realizar una revisión de otras variables socioeconómicas que favorezcan establecer acciones de apoyo por parte de las autoridades estatales o escolares.
3. Se debe insistir en la certificación de los docentes y comprobar el cambio de tendencia que se tiene en el coeficiente del modelo, una vez que haya una mayor proporción de casos certificados en cada escuela.
4. Entre las variables potencialmente eficaces para el desempeño escolar está la proporción de escuelas sin computadoras, el coeficiente en escuelas públicas es negativo por la misma causa señalada en el caso de la proporción de docentes certificados. La base de datos regional indica que hay una mayor proporción de escuelas públicas sin computadoras (produciendo coeficiente negativo), contra el caso contrario en las escuelas privadas (produciendo un coeficiente positivo). Se deberá reforzar la inversión de equipo informático para fines académicos y revertir la tendencia obtenida en las variables exploradas.
5. Efectuar un análisis que verifique si hay una asociación entre el índice de computadoras para fines académicos de la base de datos de PISA y la proporción de escuelas sin computadoras reportada por el INEE en su reporte anual, Panorama Educativo de México.

3.8 Modelo de la dimensión de pertinencia y relevancia

3.8.1 Pregunta de investigación

En el cuestionario contextual de PISA se exploran algunos aspectos relacionados con las funciones del docente y las actividades que realizan los estudiantes para el aprendizaje de Ciencias. De estos aspectos se desprenden estas preguntas de investigación:

¿Hay situaciones del proceso de enseñanza-aprendizaje suficientemente relevantes en el estudiante como para incidir en su desempeño en la prueba de Ciencias de PISA?

¿Las diferentes formas de abordar el proceso de enseñanza-aprendizaje de las ciencias son significativamente útiles?

3.8.2 Hipótesis

Las variables relacionadas con aspectos del proceso de enseñanza-aprendizaje son relevantes para el estudiante y producen diferencias estadísticamente significativas en su desempeño en Ciencias de la prueba PISA.

Todas las formas de enseñar y de propiciar el aprendizaje de la Ciencias son igualmente útiles para mejorar el desempeño de los estudiantes.

3.8.3 Variables explicativas

La dimensión de relevancia se identificó en dos componentes del análisis factorial, incluyendo variables de los niveles 1 y 2.

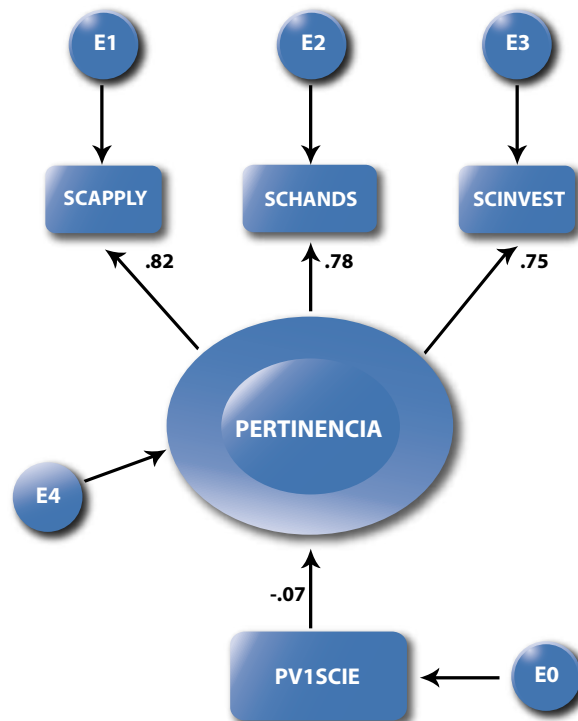
Tabla 3.32 Variables del modelo de pertinencia y relevancia

Nivel	Variables	Descripción
Variables de nivel 1	SCAPPLY	Enseñanza de las ciencias a través de aplicaciones o modelos realizados por el docente
	SCHANDS	Enseñanza de la ciencia con actividades manuales
	SCINVEST	Enseñanza de la ciencia basada en trabajos de investigaciones
Variables de nivel 2	ENVLEARN	Actividades en la escuela para aprender temas sobre el medio ambiente
	SCIPROM	Actividades de la escuela para promover el aprendizaje de las ciencias entre los estudiantes

3.8.4 Resultados del análisis factorial confirmatorio

Se construyeron dos posibles agrupamientos en los niveles 1 y 2 para representar la variable mediadora RELEVANCIA y PERTINENCIA del Modelo de Calidad Educativa del INEE y su efecto en el desempeño de la prueba PISA. Para el componente factorial del nivel 1 se tienen tres variables relativas al proceso de enseñanza-aprendizaje de las Ciencias, por medio de investigación, aplicaciones realizadas por el docente o actividades que deben realizar directamente los estudiantes. Las tres variables tienen coeficientes muy similares respecto de la variable mediadora (0.75 a 0.82), pero inciden en forma negativa en el coeficiente de regresión estandarizado de desempeño (-0.07). El ajuste es aceptable (AGFI=0.873).

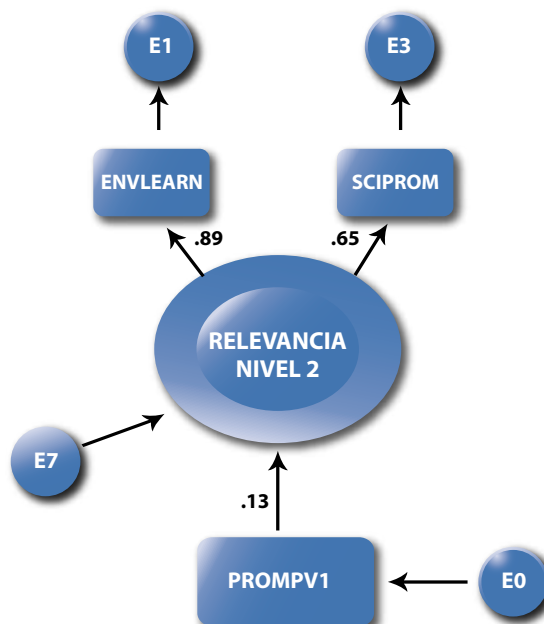
Figura 3.21 Modelo estructural de la dimensión de pertinencia y relevancia con variables de nivel I



AGFI=0.873

Para el nivel 2 se tienen dos variables en un componente factorial que se confirma con el modelo de ecuaciones estructurales (AGFI=1.0).

Figura 3.22 Modelo estructural de la dimensión de pertinencia y relevancia con variables de nivel 2



AGFI=1.0

3.8.5 Ecuaciones del modelo multinivel

La ecuación combinada del modelo jerárquico lineal que se obtiene con las variables de los dos niveles es:

$$PVSCIE_{ijk} = \gamma_{000} + \gamma_{010} SCIPROM_{jk} + \gamma_{020} ENVLEARN_{jk} + \gamma_{100} SCINVEST_{ijk} + \gamma_{200} SCAPPLY_{ijk} + \gamma_{300} SCHANDS_{ijk} + r_{0jk} + u_{00k} + e_{ijk}$$

Tabla 3.33 Resultados del modelo de la dimensión de pertinencia y relevancia

Efecto fijo		Coeficiente	Error estándar	t	Grados de libertad	p-valor
Interceptos						
Puntuación base en Ciencias	γ_{000}	424.373	7.90	53.75	6	0.000
SCIPROM	γ_{010}	24.604	4.88	5.04	1085	0.000
ENVLEARN	γ_{020}	-1.040	2.37	-0.44	242	0.660
Pendientes						
SCINVEST	γ_{100}	-13.941	1.46	-9.58	62	0.000
SCAPPLY	γ_{200}	10.566	1.70	6.20	808	0.000
SCHANDS	γ_{300}	-1.947	1.08	-1.81	39	0.078

Efecto aleatorio			Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
Nivel	Unidad							
1	Estudiante	e_{ijk}	57.48	3303.98				56.1
2	Escuela	r_{ojk}	48.17	2319.96	1079	19999.10	0.000	39.4
3	Región	u_{ook}	16.18	261.88	6	138.61	0.000	4.4

No.	Nivel		Componente de varianza en el modelo nulo	Componente de varianza en este modelo	Diferencia de varianza %
1	Estudiante	e_{ijk}	3416.98	3303.98	3.3
2	Escuela	r_{ojk}	3063.17	2319.96	24.3
3	Región	u_{ook}	441.09	261.88	40.6

Desajuste (Promedio)	335400.46	Mejora respecto del modelo de referencia	0.38%	
Diferencia respecto del modelo nulo	1264.67	Diferencia en grados de libertad	5	Diferencia significativa

Este modelo mejora al modelo nulo en más de 0.38% del desajuste promedio, permitiendo explicar 24.3% de la varianza en el nivel 2 y 40.6% en el nivel 3. En el nivel 1 se tiene 3.3% de explicación de la varianza con la inclusión de las variables de pertinencia y relevancia. Hay dos variables no significativas en sus coeficientes: ENVLEARN y el aprendizaje por medio de prácticas de laboratorio (SCHANDS) cuyo coeficiente además de ser pequeño es negativo.

3.8.6 Presentación e interpretación de resultados

El efecto más grande en el desempeño de Ciencias de PISA se tiene con la promoción de las Ciencias (tabla 3.34), con una diferencia de casi 100 puntos en la escala. Debe anotarse que simplemente dos de las variables relativas al proceso de enseñanza-aprendizaje pueden incidir (en valor absoluto) en más de 112 puntos (SCINVEST y SCAPPLY), pero el efecto de la enseñanza por investigación es negativo, lo cual es inesperado. Es el mismo caso de la variable SCHANDS aunque se confirma que su efecto es muy pequeño y, por lo tanto, no estadísticamente significativo.

Tabla 3.34 Valores de d^* para las variables del modelo de pertinencia y relevancia

Variable explicativa	Coeficiente	Categoría para resultado inferior	Categoría resultado superior	Diferencia y Variable	d^*	Significancia y clasificación
SCIPROM	24.604	-2.27	1.64	96.20	6.35	Grande
ENVLEARN	-1.040	-2.27	1.39	-3.81	-0.25	Muy pequeña
SCINVEST	-13.941	-1.26	3.03	-59.81	-3.95	Mediana
SCAPPLY	10.566	-2.46	2.63	53.78	3.55	Mediana
SCHANDS	-1.947	-2.1	2.91	-9.75	-0.64	Muy pequeña
Total teórico (valor absoluto)				223.35	14.75	Grande
Equivalente en desviaciones estándar de Ciencias				2.95		
Equivalente respecto de la escala de puntuaciones en Ciencias %				37.23		

El conjunto de variables puede producir una diferencia entre valores extremos de 223.45 puntos de la escala PISA (que es 37.23% del rango total de puntuaciones), se trata de una diferencia $d^*=14.75$ que cae en la categoría de *grande*.

3.8.7 Conclusiones e implicaciones de los resultados

De las tres variables del nivel I, la que tiene una pendiente positiva es SCAPPLY. Esta variable establece que los estudiantes reciben una enseñanza con base en aplicaciones, planteadas u organizadas por el docente, y esto redundaría en un incremento del desempeño obtenido en la prueba PISA.

Las otras dos variables del nivel I se comportan en forma inversa en sus pendientes (SCINVEST y SCHANDS), lo cual puede corresponder con un problema educativo o con una falla en la respuesta de los estudiantes a las preguntas contenidas en el cuestionario de contexto. Para la primera interpretación, supóngase que se promueve el aprendizaje de las Ciencias con actividades de laboratorio o por medio de investigaciones que deben desarrollar los estudiantes, se trata de modalidades que promueven un aprendizaje creativo centrado en el alumno, pero los coeficientes negativos obtenidos indican resultados contrarios a lo que se podría esperar. ¿Es inconveniente entonces el aprendizaje que logran los alumnos? ¿Se está orientando mal el aprendizaje? Y ¿Los docentes no promueven correctamente las actividades que deben realizar los estudiantes? Es interesante observar que el resultado es contrario al de la variable SCAPPLY que, como ya se indicó, proporciona una pendiente positiva, manifestando que el trabajo con aplicaciones, pero realizado por el docente, aparentemente ofrece buenos resultados en el desempeño de los estudiantes en la prueba PISA. Entonces ¿debe concluirse que es mejor el esquema de enseñanza centrado en las actividades del docente?

Antes de responder a estas conjeturas, debe reconocerse que el comportamiento inverso de las variables también puede estar asociado con dos aspectos, a demostrar en futuras investigaciones: el primero es que las preguntas no fueron plenamente comprendidas por los escolares y el segundo está ligado al propósito de los estudiantes al responder el cuestionario. Respecto al primer aspecto, parece que es más comprensible para el estudiante calificar lo que hace el docente que lo que realiza él mismo por su aprendizaje: debe anotarse que las preguntas que conforman a SCAPPLY (de coeficiente positivo) piden que el estudiante califique las actividades de aplicación de las Ciencias que realiza el docente en clase, en cambio las preguntas que integran a SCINVEST y SCHANDS se enfocan a la intervención del estudiante en su propio aprendizaje a través de actividades de laboratorio y de investigación que él debe efectuar directamente. Con el segundo aspecto, es posible que haya un propósito de los estudiantes por querer mostrar que tienen buenas prácticas educativas, independientemente de los resultados obtenidos en la prueba.

Las técnicas de enseñanza-aprendizaje (por medio de investigación, aplicaciones u otras modalidades), presentan un comportamiento poco consistente con lo esperado, con coeficientes positivos y negativos que no tienen el mismo efecto en el desempeño de los estudiantes, de hecho la enseñanza por investigación parece funcionar en

forma inversa a lo esperado dentro del modelo multinivel, con lo cual se rechaza la hipótesis planteada a este respecto, que consideraba que las actividades de aprendizaje incidían positivamente en el desempeño de los estudiantes.

Los resultados en estas variables dan lugar a varias conjeturas:

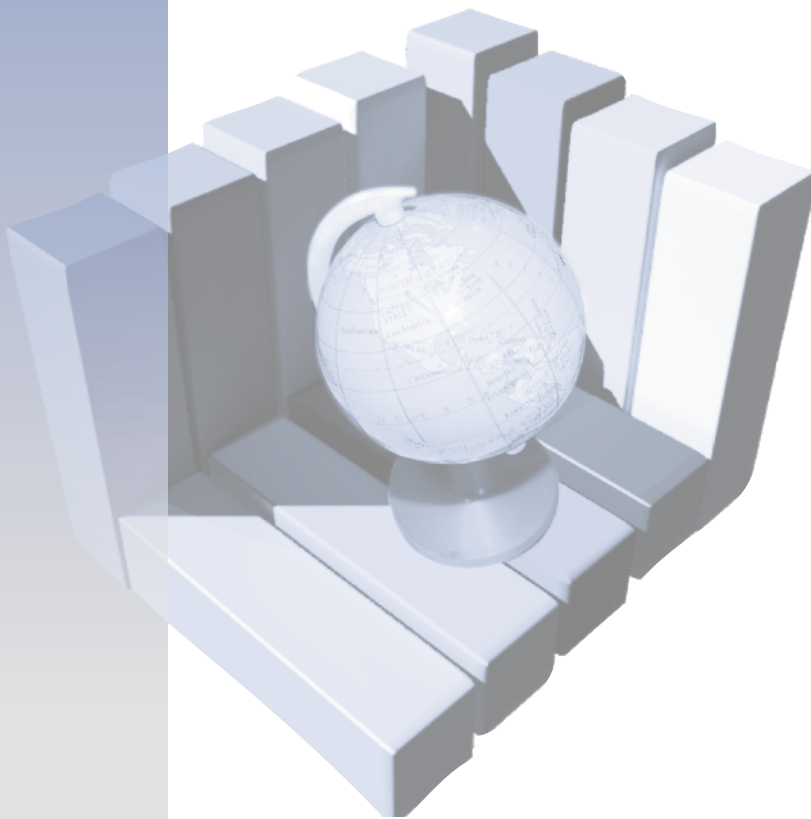
1. Los estudiantes no comprenden el sentido de las preguntas cuando se trata de actividades que deben realizar directamente, especialmente en el caso de investigación y actividades de laboratorio, porque se refieren a un modelo educativo con base en la realización de proyectos, mismo que no estaba en operación en México en la fecha de aplicación de la prueba PISA.
2. Los estudiantes comprenden mejor las preguntas en el sentido de evaluar la participación del docente y no les resulta clara la forma de responder en las preguntas relativas a actividades de aprendizaje independiente. Esto se confirma porque las preguntas de la variable SCAPPLY se refieren a lo que realiza el docente para promover el aprendizaje de las Ciencias.
3. Los estudiantes responden de la forma *socialmente deseable*, pero que no implica una relación con el desempeño real en la prueba, como ha sido observado en otros estudios similares (Díaz, Flores y Martínez, 2007).

Sugerencias

Varios puntos deberán revisarse y corroborarse en futuras investigaciones, estudiando las conjeturas de modo más detallado, en particular:

1. Corroborar que los estudiantes no comprenden el sentido de las preguntas de contexto cuando se refieren a modelos educativos en los que no han trabajado o cuando se trata de actividades o procesos de aprendizaje individual. Esto requerirá una validación de las preguntas y su replanteamiento, así como el empleo de preguntas que permitan verificar qué modelo de enseñanza-aprendizaje se desarrolla en la escuela y en el aula, para poder comparar contra las respuestas de los estudiantes.
2. Complementar un estudio que demuestre o refute que hay respuestas *socialmente deseables*, no asociadas con el desempeño del estudiante.
3. Verificar el ajuste al modelo de Rasch de las preguntas que intervinieron en las tres variables del nivel I y validar la construcción de las variables a partir de las preguntas disponibles en el cuestionario de contexto de PISA.

Conclusiones



Conclusiones

Comentarios generales

El estudio que se ha presentado aquí, se realizó con datos procedentes de la prueba PISA 2006 para verificar e interpretar el Modelo de Calidad Educativa del INEE. La base de datos está integrada por índices que se obtuvieron con las respuestas de los estudiantes y de las autoridades de la escuela a las preguntas planteadas en los cuestionarios contextuales; también incluyen las puntuaciones obtenidas por los estudiantes en la prueba de Ciencias de PISA. Estas bases se complementaron con un juego de datos socioeconómicos, demográficos y académicos a nivel región.

El marco teórico del trabajo define las dimensiones de calidad, organiza las variables de cada dimensión y sirve de base para construir modelos jerárquicos lineales en los tres niveles de anidamiento considerados en el estudio. El considerar como marco teórico al Modelo de Calidad Educativa del INEE es uno de los aspectos más atractivos del estudio, porque mostró ser un modelo aprovechable a pesar de que el diseño de la base de datos de PISA tiene su origen en un modelo conceptual diferente; tampoco se tuvieron limitaciones al complementarla con información regional disponible en trabajos previos del INEE (PIB, IDH, marginación, entre muchas otras variables).

La metodología seguida en este estudio de investigación *ex post facto*, parte de preguntas de investigación e hipótesis para cada una de las dimensiones definidas con las variables englobadas en los componentes obtenidos con un análisis factorial exploratorio. Para cada dimensión se efectuaron análisis factoriales confirmatorios con ayuda de modelos de ecuaciones estructurales; la dimensión en estudio se concibe como variable mediadora latente que asocia a las variables del componente factorial (variables independientes) con el desempeño en Ciencias de la prueba PISA (variable dependiente). El modelo de la dimensión se traduce finalmente en una expresión jerárquica multinivel proporcionando dos tipos de resultados: (1) los coeficientes de las variables de la dimensión en estudio, que participan sobre el desempeño de los estudiantes en la prueba y (2) factores de participación que tienen los niveles de anidamiento en la varianza total de las puntuaciones en Ciencias (parámetros aleatorios del modelo)¹¹.

¹¹ Los coeficientes que se obtienen con el análisis multinivel son únicos para cada variable explicativa, constantes para todas las unidades de la base de datos. Este enfoque difiere de los estudios que

A partir de los resultados se realizaron algunas propuestas de utilidad para las autoridades estatales y nacionales, los directivos y autoridades escolares, los docentes, las familias y el estudiante mismo. Dicha información corrobora la mayoría de las hipótesis propuestas para el estudio, comprobándose la influencia de algunas de las variables personales, escolares y regionales en el desempeño en Ciencias. En este momento, seguramente se trata de elementos para la reflexión que pueden servir como base para estudios posteriores.

La tabla 3.35 reúne resultados de los modelos presentados en este estudio, en particular se trata de la diferencia porcentual de la varianza obtenida en cada modelo respecto del modelo incondicional de referencia (también conocido como modelo nulo o vacío), donde no se tienen variables explicativas. Al incluirse variables en cada modelo sobre una de las dimensiones del Modelo de Calidad, se reduce la varianza en cada nivel, por ello los porcentajes de la tabla se interpretan como la proporción de varianza que explican las variables del modelo en cada uno de los niveles.

De acuerdo con los resultados obtenidos, en el nivel 1 (donde las unidades son los estudiantes) los modelos que tienen una menor varianza que la del modelo nulo, son los de la dimensión de impacto (de hecho, *Impacto 1* explica 10.6% de varianza con respecto al modelo incondicional). En el nivel 2 (donde las unidades son las escuelas) los modelos de la dimensión de equidad son los que producen una mayor explicación de la varianza (*Equidad 1* reporta 66% de varianza explicada). Finalmente, en el nivel 3 (donde las unidades son las regiones) las variables del modelo de la dimensión de eficacia explican casi la totalidad de la varianza (99.9% de varianza explicada respecto del modelo incondicional)

Tabla 3.35 Porcentajes de varianza explicada por los modelos en cada nivel

	Diferencia porcentual respecto del modelo incondicional		
	Nivel Estudiante	Nivel Escuela	Nivel Región
Situación del estudiante	2.4	35.8	62.0
Equidad 1	1.0	66.0	94.4
Equidad 2	1.2	46.2	71.3
Eficiencia	0.4	32.8	68.8
Impacto 1	10.6	19.7	23.5
Impacto 2	6.5	11.8	17.7
Eficacia 1	0.0	0.4	99.9
Eficacia 2	0.1	29.3	99.8
Pertinencia y relevancia	3.3	24.3	40.6

Queda un porcentaje alto de varianza en el nivel 1 que no es explicada por ninguna de las dimensiones ni de las variables elegidas. Esto es consecuencia de la

se realizan por medio de regresiones sobre subconjuntos de datos por nivel o por unidad (escuela, estado, región) y que producen coeficientes diferentes para cada caso.

gran dispersión que tienen los resultados en Ciencias de los estudiantes (ver por ejemplo la Figura 3.7).

Para la realización del estudio se diseñaron y revisaron dos modelos adicionales, uno para el estatus socioeconómico (ESCS) y otro para la posesión de bienes en el hogar (HOMEPOS), debido a que se trata de índices que engloban a otras variables de la base de datos de PISA. También se hizo un análisis exploratorio, con ayuda del *software* HLM, para encontrar variables regionales (de nivel 3) cuya participación fuera estadísticamente significativa en el modelo multinivel. Al presentar el modelo de eficacia se comentó que las variables regionales están altamente correlacionadas y que la fracción que representa el nivel 3 en la varianza total es muy pequeña.

Para juzgar el efecto que tiene cada variable dentro de cada una de las dimensiones de calidad se utilizó el concepto del efecto de tamaño, establecido por el coeficiente adimensional d de Cohen, que identifica diferencias significativas a partir de 0.2 (por arriba de este umbral se trata de diferencias que pueden clasificarse como pequeñas, medianas o grandes, mientras que por debajo de 0.2 se consideran no significativas, por ser muy pequeñas). Al normalizar este coeficiente dividiéndolo entre el umbral de referencia de 0.2 se obtiene el coeficiente d^* que se compara muy fácilmente contra 1 para establecer diferencias significativas. La ventaja del coeficiente d^* (o también d^*) es que no se ve afectado por el número de sujetos en la muestra, como es el caso de otros estadísticos utilizados generalmente en pruebas de hipótesis. Con el uso de d^* se dictaminan las diferencias producidas en el desempeño en Ciencias por las categorías extremas de cada variable, con lo cual se obtiene un conjunto de variables que participan de manera preponderante en los resultados de los estudiantes y otras que pueden representar áreas de oportunidad para proponer acciones de apoyo estatales o nacionales, esquemas de mediación psicopedagógica en el aula y en la escuela, así como recomendaciones para la familia y el estudiante. Todas estas acciones pretenden incidir en una mejora de las puntuaciones del desempeño de los estudiantes en Ciencias de la prueba PISA pero, más que los resultados en la prueba, lo que se intenta es promover y apoyar el desarrollo de competencias científicas que pueden ser de utilidad para la vida y el desarrollo personal.

Se puede formular una pregunta para cada uno de los niveles explorados: ¿Qué pueden hacer el estudiante y su familia, el docente en el aula y los directivos de la escuela, las autoridades estatales y nacionales, para mejorar el nivel y el desarrollo de competencias en Ciencias? Pueden prepararse algunas respuestas a esta pregunta con apoyo en los resultados obtenidos en este estudio.

Sugerencias a nivel del estudiante y de su familia

El estudiante y su familia pueden reforzar los aspectos socioculturales que influyen en el desarrollo de la persona, en especial apoyando más a las mujeres para que se involucren en temas científicos y que reconozcan el papel que la mujer ha representado a lo largo de la historia y dentro de la moderna organización social.

La familia y el estudiante, dentro de sus posibilidades económicas y sus intereses, deberán allegarse de bienes que aporten beneficios respecto de las competencias

para la vida; se trata de reducir el consumo, en caso de que así lo consideren pertinente, de productos que no inciden en una mejora sociocultural y cambiar las preferencias hacia la consulta en libros y otras fuentes de documentación, la asistencia a eventos culturales y museos. Debe resaltarse que no todos los bienes que se conoce como culturales involucran un alto costo, en realidad se trata de elegir programas culturales de televisión o de radio, en lugar de otros que no inciden positivamente en el desarrollo de la persona; el estudiante debe interesarse por la lectura de periódicos y revistas sobre temas de mayor fondo cultural; simplemente el entorno familiar puede favorecer el respeto por la naturaleza mediante acciones que impliquen actitudes y valores en pro de la ecología y del cuidado ambiental.

Sugerencias a nivel de la escuela

El docente en el aula y los directivos de la institución educativa deben tomar en cuenta que el rezago de los estudiantes y el hecho de ser mujer inciden en una diferencia desfavorable de 15 puntos en la escala de Ciencias. Se sugiere que se establezcan esquemas de apoyo a los estudiantes para que no se encuentren en situación de rezago, es decir, en caso de ser posible, lograr que los estudiantes de 15 años se encuentren en educación media superior en lugar de secundaria, atendiendo a que las condiciones académicas y sociales del nivel medio superior brindan un entorno favorable para el desarrollo personal y de las competencias del estudiante; asimismo se deberá trabajar con esquemas de intervención motivacional, de autoestima y de desarrollo de los estudiantes con rezago. También se deberá lograr que el ambiente escolar apoye y promueva la participación de la mujer en las asignaturas científicas.

Otras variables de efecto importante a nivel de la escuela son la proporción de computadoras para uso académico en la escuela y la intervención de los directivos en la definición de los modelos curriculares. En relación con esta última variable, los directivos deberán buscar esquemas apropiados para sus estudiantes según las características de la escuela y el nivel socioeconómico (pública o privada, urbana o rural, marginalidad, nivel académico); no se trata de cambiar el programa de las asignaturas, sino incluir esquemas de promoción y mejora de las competencias en temas científicos, apoyando a los estudiantes más desfavorecidos. El número de estudiantes en la clase no parece ser un factor decisivo en la mejora educativa, por ello debe adaptarse el modelo de enseñanza-aprendizaje que se emplea en la institución, propiciando esquemas reflexivos y de desarrollo de competencias creativas en el estudiante. El enfoque ambientalista, plasmado en cuestiones ecológicas y aspectos de desarrollo sustentable, debe promoverse por las autoridades escolares y directamente por el docente en el aula; deben incluirse conocimientos y habilidades cognitivas junto con el desarrollo de cualidades actitudinales, todo lo cual parece incidir en un mayor desempeño en Ciencias. Se deben revisar los enfoques de aproximación a las Ciencias dentro del aula, respaldados por políticas institucionales, cuidando la forma de proveer a los estudiantes de información sobre las carreras científicas y sobre las opciones de trabajo en esta área profesional. Una mejora de la aprecia-

ción de las Ciencias, sus aplicaciones y uso en la vida cotidiana, puede incrementar el desempeño en este tipo de asignaturas. Las autoridades de la institución deben promover y reforzar la certificación de los docentes; pero esta certificación no es suficiente si no se manifiesta en una mejora en el aprendizaje de los estudiantes; el proceso de certificación debe acompañarse de acciones que permitan comprobar que se produce una tendencia positiva para incrementar el desempeño en Ciencias de los estudiantes.

Sugerencias a nivel estatal y nacional

Las autoridades estatales y nacionales deben reconocer que pueden establecerse medidas de apoyo que incidan en algunas de las características del sistema educativo que marcan diferencias entre los estudiantes:

Por ejemplo, las diferencias en favor de las escuelas privadas sobre las públicas y de las escuelas urbanas sobre las rurales es de más de 30 puntos, por lo que es conveniente reducir la brecha académica entre estos tipos de escuelas y establecer acciones enfocadas a apoyar a los estudiantes de menores recursos, especialmente en los estados con mayor nivel de marginación.

El estatus socioeconómico del estudiante es una de las variables de mayor efecto, especialmente cuando interviene en el término cuadrático; esta variable por sí sola puede producir una diferencia de 256.5 puntos entre valores socioeconómicos extremos de una región dada; el estatus socioeconómico de la escuela también tiene un efecto importante en las puntuaciones de Ciencias. La no linealidad de las variables socioeconómicas obliga a las autoridades responsables de políticas de apoyo económico a reforzar más a las personas de niveles socioeconómicos inferiores y hacer grandes inversiones para obtener resultados que representen una ganancia significativa; el análisis muestra que no se debe caer en esquemas que sólo favorezcan a los niveles socioeconómicos altos (por ser donde se obtienen mayores resultados con menor inversión) sino reforzar a los menos favorecidos para reducir la distancia en las competencias entre grupos socioeconómicos extremos; esto se sustenta también en que las variables sociales, económicas y académicas del nivel 3 participan en la dimensión de eficacia del Modelo de Calidad para mejorar el desempeño de los estudiantes.

Las políticas educativas estatales deberán promover el desarrollo de competencias para la vida en todos los niveles, con énfasis desde el nivel de secundaria. Los docentes deberán certificarse y estar preparados para utilizar esquemas de enseñanza-aprendizaje que incidan en la comprensión de fenómenos naturales y científicos, para resolver problemas en esta área de conocimiento.

Se observó que las respuestas a las preguntas del cuestionario de contexto sobre los procesos de enseñanza-aprendizaje centrados en investigación o actividades de laboratorio que deben hacer los mismos estudiantes, difieren en su tendencia respecto de las preguntas sobre procesos centrados en las labores del docente. Se deberá verificar si se trata de una deficiencia de comprensión de las preguntas del cuestionario contextual de PISA o defectos propios del proceso educativo en clase.

Futuras líneas de investigación

Los numerosos hallazgos de este estudio no son más que un primer esbozo de la riqueza de información que se tiene con las bases de datos del proyecto PISA, sobre las cuales se podrán emprender otros trabajos de análisis y el desarrollo de líneas de investigación cualitativa y cuantitativa, no solamente en Ciencias sino también en las pruebas de Lectura y Matemáticas. No debe soslayarse la necesidad e importancia de realizar más estudios de regresión, explicativos, estructurales, multinivel o de otra índole, siguiendo la metodología de este estudio u otra conveniente para los fines de nuevos proyectos. Lo importante es reforzar las áreas de investigación para que trabajen sobre los datos de la prueba PISA y promuevan que investigadores de diversos ambientes utilicen las bases de datos, produzcan más estudios y análisis que aporten ideas de mejora al Sistema Educativo Nacional. Se deberá procurar que se realicen más estudios comparativos entre México y otros países participantes en el proyecto PISA, –de manera especial con los del Grupo Iberoamericano de PISA (GIP)– por nivel escolar, tipo de escuela u otra característica de las escuelas o de los estudiantes.

Se sugiere realizar estudios complementarios para confirmar las dimensiones de calidad del modelo propuesto por el INEE, las variables que intervienen en cada caso y la posibilidad de establecer descripciones más completas de cada dimensión o constructo, con una validación cuantitativa (utilizando ecuaciones estructurales) y por procedimientos donde intervengan jueces expertos. El estudio actual se construyó con la puntuación general en la prueba de Ciencias, así que una extensión inmediata estriba en analizar cada una de las subáreas de Ciencias y también su relación con cada nivel de desempeño (de acuerdo con la clasificación de niveles de la prueba PISA) y con los grupos extremos, especialmente porque en México hay grupos por debajo de puntuaciones esperadas y, en contraste, estudiantes con puntuaciones muy altas, aunque inferiores a las de los jóvenes de otros países. En próximos análisis se pueden estudiar los grupos de estatus socioeconómico y cultural de *élite*, en comparación con los estudiantes que están en los niveles 0 y I de PISA, para establecer la diferencia de desempeño.

La base de diseño de la prueba PISA considera que la población objetivo son los estudiantes de 15 años, cuyas competencias en Ciencias se exploran con preguntas que, en principio, no se asocian con contenidos académicos alineados a un currículo específico; se espera que las competencias provengan del aporte de diversas fuentes, medios de información y recursos propios de la edad y madurez o etapa de desarrollo de los jóvenes. Sin embargo, los resultados generan la conjetura de que la diferencia de 45 puntos entre los estudiantes de educación media superior y los de secundaria se deba al proceso académico. Se puede considerar un proyecto de investigación para medir la posible influencia del nivel escolar en los resultados de PISA y descartar o confirmar, en su caso, si las competencias están asociadas con una intervención académica.

Se contó en la base de datos con los resultados de la prueba Excale en Matemáticas o en Lenguaje, pero los datos no aparecieron de forma clara en ninguno de los

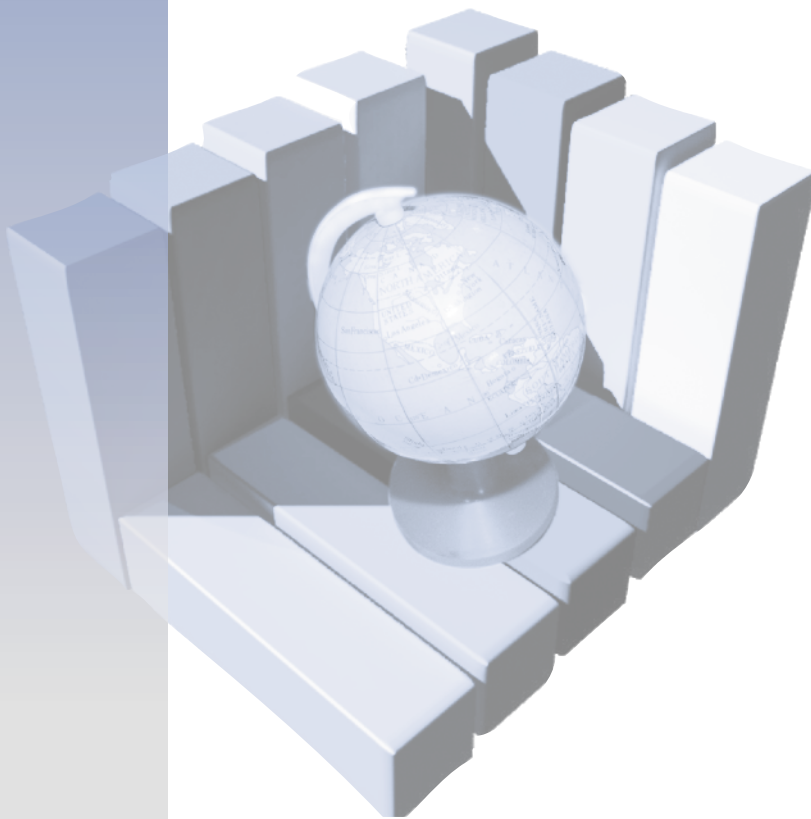
componentes factoriales, lo cual no permite verificar en este momento si los resultados de la prueba del INEE pueden servir como variables explicativas de los resultados de Ciencias de la prueba PISA. Un proyecto a desarrollar en el futuro deberá centrarse en la asociación de resultados de las áreas semejantes que tienen ambas pruebas (Lectura y Matemáticas) y en el análisis de la relación de ellas con la permanencia y la continuidad de los estudiantes en el sistema escolar.

Para el presente estudio se corroboró la calibración de los ítems y la validez de la escala PISA utilizando la distribución de ítems definida por la recta de diseño. Es imperioso realizar más estudios de validez de la prueba recabando el mayor número posible de evidencias (validez facial, de contenido, de criterio y de constructo), tanto de los ítems de los cuadernillos de prueba como de los cuestionarios de contexto. Un análisis detallado de algunos de las preguntas de contexto permitirá dilucidar si las respuestas inesperadas de los alumnos se deben a una incomprensión de lo que se pregunta, a un propósito de responder en forma *socialmente deseable* o a deficiencias dentro del proceso de enseñanza-aprendizaje. Cualquiera de estas opciones dará lugar a acciones correctivas importantes para mejorar las preguntas y posiblemente elevar el nivel de desempeño de los estudiantes en la prueba PISA, no como resultado de una preparación ante la prueba sino como una mejor preparación ante la vida misma.

Es conveniente complementar el cuestionario de contexto con objeto de contar con un mayor número de variables que permitan comparar algunos patrones de consumo del estudiante y juzgar la pertinencia de las variables empleadas en el índice de posesión de bienes culturales y de otros índices potencialmente asociados con el desempeño de los estudiantes. Con esta información se podrán proponer estrategias promocionales para la adquisición de libros y otros materiales educativos; otorgamiento de becas y créditos u otras modalidades de apoyo al estudiante y su entorno familiar.

Un análisis de la medida y del ajuste al modelo de Rasch de los ítems del cuestionario de contexto, junto con el estudio del funcionamiento diferencial de los ítems, brinda una línea de trabajo que puede ser benéfica para analizar las variables que tienen comportamiento anormal o inesperado. Desde un punto de vista más técnico, se sugiere contar con la participación de la Dirección de Indicadores educativos del INEE para verificar y dominar la metodología de diseño de los índices a partir de ítems del cuestionario de contexto.

Referencias



Referencias

- Adams, R. y Wu, M.L. (2002) *PISA 2000 Technical Report*. Paris: OECD Publishing.
- Aguayo, S. (2008) *El almanaque mexicano*. México: Editorial Santillana. 287 pp.
- Allison P. (2002) *Missing Data*. SAGE University paper, no. 136. Series: quantitative applications in the social sciences. California: Sage University. pp. 9-10.
- American Psychological Association (1994) *Publication manual of the American Psychological Association*, 4a ed. Washington D.C.: APA. pp. 11.
- American Psychological Association (2001) *Publication manual of the American Psychological Association*, 5a ed. Washington D.C.: APA.
- Arbuckle, J. (2003) *AMOS 5.0*. Software. Philadelphia: Amos development Corp.
- Arbuckle, J. y Wothke, W. (1999) *AMOS 4.0 User's Guide*. Chicago: SmallWaters. pp. 452
- Backhoff, E.E., Bouzas, R.A., Contreras, C., Hernández, P.E. y García, P.M. (2007a) *Factores escolares y aprendizaje en México*. El caso de la educación básica. México: INEE. pp. 168.
- Backhoff, E.E., Bouzas, R.A., Hernández, P.E. y García, P.M. (2007b) *Aprendizaje y desigualdad social en México*. Implicaciones de política educativa en el nivel básico. México: INEE. pp. 151.
- Backhoff, E.E., Díaz, G. M.A. (2005) *Plan General de Evaluación del Aprendizaje. Proyectos Nacionales e Internacionales*. INEE. México. Pág.52
- Bryk, S. A. y Raudenbush, S. W. (1992) *Hierarchical linear models. Applications and data analysis methods*. London: Sage Publications. pp. 265.
- Bryk, S., Raudenbush, A. y Congdon, R. (2005) *HLM Versión 6.03. User's Guide to the Software*. London: Sage Publications.
- Byrne, B.M. (2001) *Structural equation modeling with AMOS*. Multivariate application series. New Jersey: Lawrence Erlbaum Associates. pp. 338.
- Cansino Muñoz-Repiso, J. (2001) *La Eficiencia del Sector Público: Métodos de Evaluación y Organismos Responsables*. El Caso de España. Madrid: Instituto de Estudios Fiscales.
- Centra, J. y Potter, D. (1980) *Schools and teacher effects: an international model*. Review of Educational Research. pp. 50, 273-291.
- Chinen, L.M. (2006) *Análisis de los resultados de la Prueba Nacional de Aprovechamiento en Lectura en Secundaria: Estudio multinivel de logro y tendencias*. INEE. México. 168 pp.

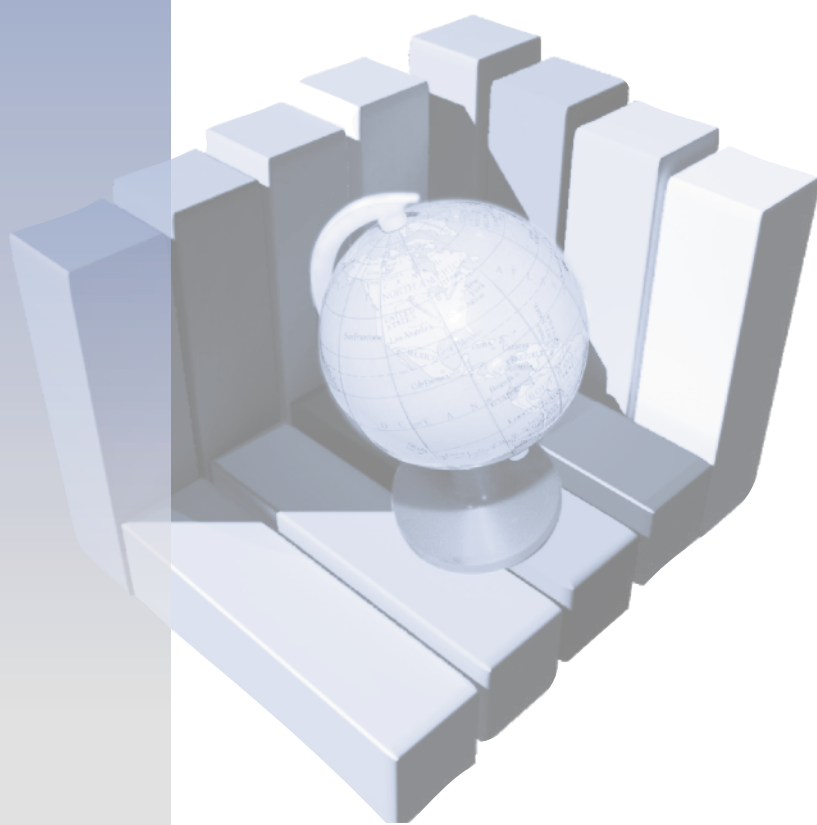
- Chu, A., Morganstein, D. y Wallace, L. (1992) *Sampling*. Part I. IAEP Technical Report.
- Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*. LEA: Hillsdale, N.J. pp. 20-40.
- Cohen, J., y Cohen, P. (1983) *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.) Hillsdale, NJ: Erlbaum.
- Coleman, J.S. et al. (1966) *Equality of Educational Opportunity*. National Center for Educational Statistics, Report Number OE-38001. U.S. Washington, D.C: Government Printing Office. pp. 737.
- Coleman, J., Hoffer T. y Kilgore S. (1981) *Public and private schools*. National Center for Education Statistics. U.S. Washington, D.C: Government Printing Office. pp. 415.
- De la Orden, H. A. (1985) *Hacia una conceptualización del producto educativo*. Revista Investigación Educativa, 3 (6), pp. 271-283.
- De la Orden, H. A., Asensio, I., Carballo, R., Fernández Díaz, J., Fuentes, A., García Ramos, J.M. y Guardia, S. (1997) *Desarrollo y validación de un modelo de calidad universitaria como base para su evaluación*. RELIEVE, vol. 3, n. 1. Consultado en http://www.uv.es/RELIEVE/v3n1/RELIEVEv3n1_2.htm en octubre 2008.
- De Landsheere G. (1976) *Introduction à la recherche en éducation*. Paris: Armand Colin-Bourrelle. pp. 22-45.
- Díaz G.M.A., Flores V.G. y Martínez R.F. (2007) *PISA 2006 en México*. México: INEE. pp. 343.
- Edmonds, R.R. (1979) *Effective schools for the urban poor*. Educational Leadership, 37 (1) 15-25.
- Else-Quest, N.M., Shibley, H.J., Hill Goldsmith, H. y Van Hulle, C.A. (2006) *Gender Differences in Temperament: A Meta-Analysis*. Psychological Bulletin, Vol. 132, No. 1, 33-72
- Gago, H.A. y Mercado del C., R. (1995) *La evaluación en la Educación Superior Mexicana*. Revista de la Educación Superior, núm. 96. México: ANUIES. pp. 61-86.
- Gaviria, J. L., Martínez, A. R. y Castro, M. (2004) *Un estudio multinivel sobre los factores de eficacia escolar en países en desarrollo. El caso de los recursos en Brasil*. Education Policy Analysis Archives, 12 (20) Recuperado el 10/18/2006 de <http://epaa.asu.edu/epaa/v12n20/>
- Goldstein, H. (1999) *Multilevel statistical models*. London: Edward Arnold. 163 pp.
- Hanushek, E.A. y Luque, J.A. (2003) *Efficiency and equity in schools around the world*. Economics of education review, 22, 481-502
- Harvey, L. y Green, D. (1993) *Defining quality*. Assessment and Evaluation in Higher Education, 18 (1), pp. 9-34.
- Hernández, S. R., Fernández, C.C. y Baptista, L.P. (2004) *Metodología de la investigación*. 3ª edición. México. McGraw-Hill.
- Hox, J.J. (1995) *Applied multilevel analysis*. Amsterdam: TT-Publikaties. pp. 118.
- Instituto Nacional para la Evaluación de la Educación (2004) *¿Qué es la calidad educativa?* Colección de folletos. Num. 3, Los temas de la evaluación, ISBN 1665-9465, pp. 10
- _____ (2006) *Plan Maestro de Desarrollo 2007-2014*. México: INEE. pp. 63.

- Kelley, T.L. (1939) *The selection of upper and lower groups for the validation of test items*. Journal of Educational Psychology. pp.17-24.
- Kerlinger, F.N. y Lee, H. (2002) *Investigación del comportamiento*. 4ª edición. México: McGraw-Hill.
- Kreft, I. y De Leeuw, J. (1998) *Introducing multilevel modeling*. London: Sage Publications.
- Larroyo, F. y Cevallos M.A. (1965) *La lógica de las ciencias*. 15ª ed. México: Ed. Porrúa. PP 255 y sigs.
- Lindsay, A.W. (1982) *Institutional performance in higher education: The efficiency dimension*. Review of Educational Research, 52 (2), pp. 175-199.
- Little, R. J. A., y Rubin, D. B. (2002) *Statistical analysis with missing data*. Hoboken, New Jersey, Wiley.
- Martínez R.F. (2003) *La calidad de la educación básica en México*. Primer informe anual. México: INEE. pp. 196.
- _____ (2004) *La calidad de la educación básica en México*. Resumen ejecutivo. INEE. pp. 31.
- _____ (2005) *La calidad de la educación básica en México, 2005*. Instituto Nacional para la Evaluación de la Educación. México: INEE. pp. 180.
- Martínez, R.F. y col. (2007) *Propuesta y experiencias para desarrollar un Sistema Nacional de Indicadores Educativos*. Instituto Nacional para la Evaluación de la Educación. México: INEE. pp. 280.
- Mc Kenzie, D. E. (1983) *Research for school improvement: An appraisal of some recent trends*. Education Researcher, 12 (4), pp. 5-17.
- McKnight, P.E., McKnight, K.M., Sidani, S. y Figueredo, A.J. (2007) *Missing data. A gentle introduction*. pp 169-170
- Morrison, D.F. (1967) *Multivariate statistical methods*. New York: McGraw Hill Book Co. pp. 338.
- Mullis, I.V.S, Dossey, J.A., Owen, E.H. y Phillips, G.W. (1993) *Guidelines for sample participation and explanation of derivation of weighted participation*. Appendix B. NAEP 1992 Mathematics report card for the nation and the states. Report N. 23-ST02, Washington: ETS y US Dept. of Education. pp. 307-315
- Murillo, T. F. J. et al. (2007) *Investigación iberoamericana sobre eficacia escolar*. Bogotá: Convenio Andrés Bello. pp.376.
- Murphy, J. (1992) *The landscape of leadership preparation: Reframing the education of school administrators*. Newbury Park, CA: Corwin Press.
- Myers, A. y Hansen, C.H. (2002) *Experimental psychology*. 5a. edición. Wadsworth, Thomson Learning. EUA. pp. 117-139.
- Nunnally, J.C. y Bernstein, I.J. (1995) *Teoría psicométrica*. 3ª ed. México: McGraw Hill. pp. 843.
- OECD-PISA (2003) *Weighting and sampling variance*. Doc: NPM(0312)14. pp. 13.
- _____ (2007) *Organization for Economic Co-operation and Development PISA 2006: Science Competencies for Tomorrow's World*, Vol. 1. Annex A8. Technical notes on multilevel regression analysis. pp. 8.
- _____ (2008) *Organization for Economic Co-operation and Development. Questionnaire indices for PISA 2006*. pp. 39.

- Pardo C., Rocha M., Avendaño B., y Barrera L. (2005) Manual de procesamiento de datos y análisis de ítems, versión 2. Segundo Estudio Regional Comparativo y Explicativo. Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), Santiago. Chile. PP. 11.
- Raudenbusch, S., Bryk, A. y Congdon, R. (2005) HLM for windows, V.6.0.3. User manual help.
- Raykov, T. y Marcoulides, G.A. (2000) A first course in structural equation modeling. Lawrence Erlbaum Associates, New Jersey. 205 pp
- Robles, V.H.V. y Martínez, R.F. (2006) *Panorama educativo de México 2006. Indicadores del Sistema Educativo Nacional*. México: Instituto Nacional para la Evaluación de la Educación. pp. 605.
- Rodriguez, M.C. (2005) *Three Options Are Optimal for Multiple-Choice Items: A Meta-Analysis of 80 Years of Research*. Educational Measurement: Issues and Practice 24(2) pp. 3-13.
- Rosenthal, R. (1979) *The file drawer problem and tolerance for null results*. Psychological Bulletin. pp. 86, 638-661.
- Rowan, B., Bossert, S.T. y Dwyer, D.C. (1983) *Research on Effective Schools: A Cautionary Note*. Education Researcher, 12 (4), pp. 25-31.
- Rubin, D. B. (1976) *Inference and missing data*. Biometrika (63), pp. 581-592.
- Ruiz de Miguel, C. y Castro M.M. (2006) *Un estudio multinivel basado en PISA 2003: factores de eficacia escolar en el área de matemáticas*. Archivos Analíticos de Políticas Educativas, 14 (29) Recuperado [julio 2008] de <http://epaa.asu.edu/epaa/v14n.29>
- Scott, R. y Walberg, H.J. (1979) *Schools alone are insufficient: a response to Edmonds*. Educational Leadership, 37 (1), pp. 25.
- Scriven, M. (1969) *An introduction to metaevaluation*. Educational Product Report, 2, 36-38.
- Seidel, T. y Shavelson, R.J. (2007) *Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results*. Review of Educational Research. December 2007, Vol. 77, No. 4, pp. 454-499.
- Stoll, L. y Fink, D. (1999) *Para cambiar nuestras escuelas*. Barcelona: Octaedro.
- Stufflebeam, D.L. y Shinkfield, A.J., (1981) *Meta-Evaluation: Concepts, Standards and Uses*. en R. Berk., ed., Educational Evaluation Methodology: The State of the Art, Johns Hopkins.
- _____ (1987) *Evaluación Sistémica: Guía Teórica y Práctica*. Temas de Educación, Paidós.
- Thompson, B. (1996) *AERA Editorial policies regarding statistical significance testing: three suggested reforms*. Educational Researcher, vol.25, n.2, pp. 26-30.
- Timperley, H. y Alton-Lee, A. (2008) *Reframing Teacher Professional Learning: An Alternative Policy Approach to Strengthening Valued Outcomes for Diverse Learners*. Review of Research in Education. Vol. 32, pp. 328-369.
- Tristán, L.A. y Vidal, U.R. (2007) *Linear model to assess the scale's validity of a test*. AERA Meeting 2007, "New Developments in Measurement Thinking", SIG-Rasch Measurement) Disponible en ERIC: DOCUMENT: ED501232

- Wayne, A. y Youngs, P. (2003) *Teacher Characteristics and Student Achievement Gains: A Review*. Review of Educational Research. 73(1), pp. 89-122.
- Webster, W. J., Mendro, R.L. y Almaguer, T.O. (1994) *Effectiveness indices: A "value added" approach to measuring school effect*. Studies in Educational Evaluation, 20, 1, 113-137.
- Willms J.D. y Smith T. (2005) *A Manual for Conducting Analyses with Data from TIMSS and PISA*. pp. 9-10.
- Willms, D. (2006) Las brechas de aprendizaje: diez preguntas de la política educativa a seguir en relación con el desempeño y la equidad en las escuelas y los sistemas educativos. Instituto de Estadística de la UNESCO. Montreal, 95 pp.
- Wu, M.L. (2004) *Plausible Values*, Rasch Measurement Transactions, 18:2 p. 976-978

Anexos



Anexos

Anexo I. La calidad como variable global

En el modelo propuesto por el INEE para definir la calidad educativa, se establecen claramente las cinco dimensiones o ejes que permiten distinguir a una educación de calidad de otra que no lo es. Esto contrasta, evidentemente, con los estudios que se enfocan a una sola de las dimensiones, siendo importante destacar que la dimensión de eficacia es la que más se ha tratado de estudiar en las referencias.

Se cuenta con numerosos estudios que proponen modelos de eficacia escolar, partiendo de los extensos trabajos de Coleman (1966, 1981), Cansino, Muñoz-Repiso (2001), Hanushek, E.A. y Luque, J.A. (2003), entre otros autores. En general, los modelos brindan un acercamiento (en forma prevista o no) a modelos de tipo CIPP, con mayor o menor éxito. En algunos casos se afirma que la eficacia descansa en variables de contexto del estudiante, de la escuela o del país; en otros casos se dice que depende del liderazgo de los directivos o del uso de los recursos disponibles, lo cual recae más apropiadamente en el concepto de la eficiencia, puesto que contempla no la cantidad de recursos sino su empleo óptimo para lograr los propósitos previstos. En este contexto, la eficiencia puede ubicarse en diversos niveles, desde el estudiante y su familia, pasando por la escuela, hasta llegar al estado o el país, lo cual propicia la toma de decisiones de política educativa. A este contexto se agregan variables de inversión nacional o estatal, como es el PIB nacional o la fracción del mismo destinado a educación en sus diversos grados.

Los modelos se han enriquecido al incluir la estructura escolar, las condiciones de contratación y la estabilidad de los docentes, así como su preparación y experiencia; de este modo, el nivel socioeconómico, apuntado por Coleman en un principio, ha evolucionado hacia modelos más comprensivos que incluyen los atributos y cualidades de los docentes (Coleman et al., 1981). Este autor, en un estudio longitudinal realizado sobre escuelas públicas y privadas, hace patente el papel de las escuelas privadas en el nivel educativo al favorecer la obtención de mejores resultados en varios aspectos, como el desempeño cognitivo, los niveles de autoestima, el ambiente de disciplina, por ejemplo.

De la Orden y col. (1997) mencionan que en la literatura se asimila la calidad educativa a aspectos de gestión, lo cual da lugar a acciones de control de calidad, auditoría, evaluación, políticas y asignación de fondos públicos o privados. En esta óptica de gestión, es interesante notar que los modelos de control de calidad pueden ser diseñados independientemente de la definición que se tenga sobre la calidad educativa. Las concepciones de calidad de Harvey y Green (1993) referidas por De la Orden y col. (1997), tratan de aclarar el concepto al establecer criterios que permiten medir y evaluar la calidad educativa. Un modelo como el propuesto por dichos autores establece relaciones entre la eficacia, la eficiencia y la funcionalidad, no solo como parte de un modelo de calidad, sino como partes ligadas al contexto sociocultural y económico de alguno de los niveles educativos (en particular el nivel universitario), el modelo se vuelve *plano*, sin facetas y con relaciones encadenadas entre variables; de este modo los *procesos* corresponden con la *eficiencia*, la *calidad* se plasma en la *eficacia*, es decir, la equivalencia entre dimensiones es por momentos confusa, incompleta o sesgada, pero tiene la ventaja de proveer las bases para la evaluación sobre indicadores cuantificables, identificados con criterios y predictores de funcionalidad, eficacia y eficiencia.

Para el caso de Iberoamérica, Murillo y col. (2007) realizó una recopilación muy interesante de estudios donde se muestra un gran conjunto de modelos explicativos de proyectos nacionales (como los de las pruebas SABER en Colombia) o internacionales realizados por el LLECE. Se trata de estudios donde se hacen intervenir factores de contexto del docente o del estudiante; socioeconómicos de la persona y de la escuela; también se incluyen elementos evolutivos, administrativos y sociales, como la asistencia a los grados escolares iniciales, además de conceptos como nutrición, marginalidad y entorno urbano-rural, la influencia de la lengua materna indígena y el bilingüismo.

El crecimiento de este tipo de estudios se vio favorecido por el uso del análisis multivariado y de los modelos de regresión multinivel, con los trabajos de Aitkin y Longford (1986) y finalmente con el desarrollo de software por parte de Bryk y Raundebush (1992, 2005). En estos análisis multivariados se pueden considerar los niveles de anidamiento y la intervención de variables de contexto muy diversas. Esto ha representado una línea de trabajo para el análisis de los factores asociados al rendimiento escolar, que se complementan con modelos de ecuaciones estructurales.

ALGUNAS DEFINICIONES ALREDEDOR DE LA CALIDAD

Continuidad de los efectos escolares. Supervivencia del material y de las habilidades aprendidas por los alumnos, después de varios años de haber salido de la institución educativa. Es el sello, la marca o la influencia que ha dejado la escuela en los alumnos con el paso de los años.

Desarrollo integral del alumno. Rendimiento académico del estudiante en áreas sustantivas, incluyendo competencias básicas (lenguaje y matemática), competencias de alto nivel (creación, capacidad crítica, solución de problemas) y competencias ciudadanas (valores, actitudes, relaciones y comunicación con otros miembros de la sociedad).

Efecto de escolarización. Proporción de altos resultados en el rendimiento escolar debido a la intervención de la escuela. Se puede comparar el efecto de escolarización entre alumnos que asisten a la escuela de los que no asisten.

Efecto escolar. Capacidad de la escuela para influir en los resultados de sus alumnos. Matemáticamente se puede expresar por medio del porcentaje de varianza del rendimiento del alumno debido a la escuela. El efecto escolar, dentro del modelo multinivel, se estima a través de la correlación intra-clase de las variables de ajuste. De esta forma, el proceso de cálculo es el siguiente: en primer término, se obtienen los modelos multinivel ajustados —uno para cada valor plausible de la variable producto— y de ahí se obtienen las correspondientes correlaciones intraclase.

Eficacia. Medida del éxito en los factores de producto del modelo CIPP. De la Orden (1985) comenta la dificultad de establecer medidas de eficacia cuando se trabaja en función de metas u objetivos, debido a la indefinición de ellas o a la deficiencia en el establecimiento del constructo de eficacia en términos de productos. Por ello el interés de operacionalizar el concepto de eficacia dentro del modelo de calidad. Véase la definición de eficacia en el modelo del INEE.

Eficacia escolar. Se refiere a la cualidad que permite a una institución educativa conseguir los fines propuestos con los medios adecuados. La eficacia escolar depende de algunas variables mediadoras que pueden influir en los resultados, especialmente de las escuelas privadas, tales como: contratación y relación laboral, la organización institucional, el tamaño de los grupos, el nivel socioeconómico de la institución, la profesionalización docente, vista como obtención de un título de licenciatura o un grado académico. La escuela de calidad es la que, además, se propone fines social y humanamente relevantes según Ruiz de Miguel y Castro (2006), pero esta definición ya no sería independiente de la dimensión de relevancia y trascendencia. Véase la definición de eficacia en el modelo del INEE.

Eficacia diferencial. Variación de los efectos escolares entre grupos de alumnos. Se dice que está relacionada con la validez externa de la medida del efecto escolar y su poder de generalización a las poblaciones participantes. También se reporta que la eficacia diferencial tiene que ver con el rendimiento previo (o nivel de entrada), con algunas discrepancias entre autores.

Eficiencia. Medida del uso de los recursos empleados para obtener los resultados; concepto de hacer lo mejor posible con los recursos de que se dispone. Depende de variables mediadoras como: la cantidad de dinero o inversión económica que se destina a educación, pudiendo ser en monto total o en porcentaje del PIB. Se acostumbran expresar por medio de relaciones entre los medios y los productos. Según Lindsay (1982) hay tres índices de eficiencia: económica, administrativa y pedagógica, según se enfoque a la productividad de los recursos financieros, la gestión de dichos recursos o la participación de los insumos docentes y psicopedagógicos. Véase la definición de eficiencia en el modelo del INEE.

Equidad. Desarrollo de cada uno de los alumnos por acciones concretas realizadas por la institución, o la obtención de buenos resultados académicos o de otro tipo definidos por el perfil de estudiante deseado. Según Murphy (1992), Stoll y Fink, (1999), se consideran buenos resultados si son logrados por todos los estudiantes, lo cual es una medida de eficacia pero también de equidad. La realidad es que los resultados de la eficacia son diferenciales (con distinciones por género, por grupos étnicos, por grupos culturales, por países, por tipo de docente en la escuela); evidentemente resulta deseable la eficacia con equidad, pero no sería aceptable decir que sólo con equidad se tiene eficacia. Véase la definición de equidad en el modelo del INEE.

Escuela eficaz. Institución educativa que promueve de forma duradera el desarrollo integral de cada uno de sus alumnos más allá de lo que sería previsible teniendo en cuenta su rendimiento inicial y la situación social, cultural y económica de sus familias. Se identifica por tres características: (1) desarrollo integral del estudiante, (2) equidad y (3) valor agregado; evidentemente, esta definición plantea un problema respecto del modelo de calidad propuesto por el INEE, por la presencia de la dimensión de equidad como parte de la variable eficacia de la escuela, en lugar de ser una dimensión independiente. En una escuela eficaz todos los alumnos logran los resultados máximos que sean posibles; si se toman en cuenta las diferencias de partida, la eficacia se vuelve una característica diferencial.

Estabilidad temporal. Confiabilidad de resultados de un año a otro. Se pueden tener tres casos de estabilidad: (1) cuando los valores son comparables entre años, lo cual corresponde con la estabilidad temporal; (2) cuando hay una diferencia significativa en favor o en contra de un año a otro, en este caso no se tiene estabilidad pero permite estimar la ganancia o pérdida comparativa entre años; (3) cuando las diferencias son aleatorias entre años, lo cual no es un resultado inútil sino que indica la volatilidad de las políticas o acciones en la consecución de los resultados o, en su caso, la no pertinencia de los instrumentos de medición para lograr reportar los propósitos que se persiguen.

Funcionalidad. Relación entre variables de entrada y variables de salida en el modelo CIPP. Se establece por medio de varios criterios de relación entre elementos del modelo. Así se tienen relaciones entre:

- Valores, expectativas y necesidades → estructura, procesos directivos y de gestión. Ejemplos: estructura de autoridad y participación en el sistema de decisiones del centro escolar; clima institucional; validez cultural, social y laboral del currículo y de los programas.
- Expectativas y necesidades que se traducen en producto y resultados. Ejemplos: adecuación en cantidad y modalidad de los graduados; vigencia de conocimientos, aptitudes y competencias adquiridos; relevancia de valores y actitudes, relevancia de la aportación científica investigadora.
- Aspiraciones, expectativas, necesidades y demandas que redundan en las metas y los objetivos educativos.

Valor agregado. Diferencia de desempeño obtenido por el alumno o grupo de alumnos, tomando en cuenta el estado inicial o rendimiento previo, considerando la situación y contexto socioeconómico y capital cultural de la familia (Webster, Mendro y Almaguer, 1994). Se puede llegar a decir que una institución es eficaz si su valor agregado es superior al previsible para un juego de valores de entrada. Se mide por medio de la diferencia del valor obtenido respecto del valor esperado para un estado inicial, debiéndose tener valores positivos y altos en la escuela eficaz.

Para efectuar el análisis de los modelos multivariados se deben definir las variables explicativas que inciden en los resultados académicos en la calidad educativa. La cantidad de variables explicativas es amplísima en los estudios publicados; finalmente cada investigador utiliza diferentes puntos de referencia y conceptos de partida en los modelos descriptivos, en las regresiones múltiples o en los estudios de factores asociados al rendimiento. Pueden distinguirse, sin embargo, varios juegos de variables que se repiten entre estudios y que se acepta que tienen una influencia en el desempeño académico o en las competencias para la vida.

Variables del estudiante: sexo, edad, nutrición, antecedentes escolares (repetición, ausentismo, deserción), tiempo dedicado a hacer tareas, hábitos de estudio, actitud ante trabajo en equipo, desarrollo de proyectos e investigaciones, hábitos de lectura, autopercepción de esfuerzo, expectativas personales.

Variables de la familia: ingreso familiar, nivel educativo de los padres, estructura familiar, horas dedicadas por la familia a apoyar el trabajo académico de los hijos, disponibilidad de espacios para el estudio en casa, recursos educativos, capital cultural y socioeconómico.

Variables del docente: nivel educativo, experiencia, procesos y actitudes hacia la actualización y perfeccionamiento o formación continua, resultados en pruebas docentes; infraestructura personal o de la institución para preparar su trabajo, expectativas hacia los alumnos, hábitos de estudio, materiales didácticos que emplea (tareas, documentos y otras ayudas), percepción de los problemas del proceso de enseñanza-aprendizaje.

Variables de la escuela: infraestructura, tipo de escuela, ubicación geográfica y sector que atiende, tamaño de los grupos, matrícula, relación personal administrativo y docente por alumno; tipo de gestión o liderazgo de los directivos, disposición de espacios diferentes a las aulas (patio de juegos, laboratorio, bibliotecas y otras áreas), infraestructura de información y documentación, participación en proyectos con fondos públicos, privados, ONG o de otro origen.

Dentro de los factores del centro escolar y que inciden en la eficacia escolar están el ambiente de la institución, el liderazgo de los directivos, las altas expectativas, los métodos psicopedagógicos y el trabajo en equipo. También intervienen la disponibilidad y gestión de los recursos humanos, económicos y materiales, así como las variables de la profesión docente (formación inicial, actualización y formación continua, estabilidad laboral y condiciones de trabajo) y las actitudes y relaciones entre estudiantes, docentes y autoridades.

Según Rowan y col. (1983), las características que promueven un mejor desempeño en las habilidades básicas y una mayor eficacia educativa son:

Entorno: se trata de que haya un ambiente escolar propicio para el aprendizaje, libre de problemas de disciplina.

Expectativas: el sistema educativo debe tener altas expectativas sobre el desempeño del alumno porque se ha demostrado que esto incide en los resultados. Debe tenerse cuidado porque este factor puede traducirse en la presencia de estereotipos y modelos de halo o de influencia, como indica Edmonds, E.E. (1979).

Integralidad: se espera enfatizar a todos los ámbitos de la escuela en las habilidades básicas. Esto quiere decir que los estudiantes obtienen desempeños aceptables en todas las asignaturas.

Sistematización educativa: conjunto de objetivos educativos bien definidos, que permiten evaluar los desempeños de los estudiantes y supervisar sus avances y logros.

Gestión y dirección: las autoridades o directivos de la escuela deben contar con cualidades de liderazgo y planeación que se comparten por el resto del personal, deben ser capaces de definir altos niveles de exigencia, tener esquemas de supervisión de los avances y proporcionar incentivos en el aprendizaje.

Edmonds (1979), Scott y Walberg (1979) incluyen los aspectos de liderazgo en los directivos; altas expectativas en los estudiantes; ambiente propicio en la escuela; énfasis en el aprendizaje de habilidades básicas (verbales y matemáticas); evaluación y retroalimentación continuas; rol de la estructura y participación de la familia, junto con las cualidades y el esfuerzo del estudiante, afectado por el sexo, el nivel socioeconómico. Centra y Potter (1980) consideran que los grupos de variables que explican la eficacia escolar, entendida como resultados de aprendizaje, son: condiciones de la escuela y del municipio, estado o región; condiciones dentro de la escuela; características y comportamiento de los docentes; características y comportamiento de los estudiantes. De esta manera, el conjunto de variables es demasiado amplio y vago, porque ¿qué se entiende por condiciones dentro de la escuela? ¿Cuáles son los comportamientos de los docentes? Otros modelos, como el de Mc Kenzie (1983) sugieren que el liderazgo, en lugar de ser una variable explicativa de la eficacia, es una dimensión por sí misma, complementaria a la eficacia y la eficiencia.

La tabla AI.1 organiza las variables descritas en los niveles: estudiante, escuela (institución y aula) y región (estado y país).

Tabla A1.1 Distribución de variables por cada nivel

Estado, región o país		
	Inversión	
	<ul style="list-style-type: none"> Inversión % PIB en educación infantil, primaria, secundaria, educación media superior: 	<ul style="list-style-type: none"> % PIB destinado a educación Grado de desarrollo del país (IDH)
	Capital socioeconómico y cultural	
	<ul style="list-style-type: none"> Nivel educativo de la población Tasas de escolarización por edades 	<ul style="list-style-type: none"> Nivel socioeconómico por región Tasas de analfabetismo
	Organización y gestión	
	<ul style="list-style-type: none"> Autonomía curricular, organizativa, de gestión de recursos y de personal 	
	Tasas de rendimiento	
	<ul style="list-style-type: none"> Tasa de abandono Porcentaje de alumnado que repite Tasa de idoneidad en la edad del alumnado 	<ul style="list-style-type: none"> Relación enseñanza pública-enseñanza privada

Escuela (institución)		
Factores escolares	Infraestructura y características de la escuela	
	<ul style="list-style-type: none"> Nivel socioeconómico Buena infraestructura de la escuela Hábitat o ubicación de la escuela Tamaño de la escuela y de los grupos Recursos materiales y económicos disponibles Utilización apropiada de las instalaciones Calidad de los recursos educativos Composición del alumnado 	<ul style="list-style-type: none"> Titularidad de la escuela Proporción de profesores con certificación pedagógica Condiciones socio-laborales del profesorado y estabilidad del profesorado Autonomía para la gestión económica y relación con la comunidad educativa Composición del profesorado Trabajo en equipo del profesorado Satisfacción del profesorado
	Organización del personal	
	<ul style="list-style-type: none"> Cooperación entre personal de distintas posiciones y funciones Realización de reuniones y consultas 	<ul style="list-style-type: none"> Trabajo y enseñanza en equipo Realización de debates y reuniones académicas
	Clima escolar	
	<ul style="list-style-type: none"> Ambiente y entorno con reglas definidas Atmósfera escolar ordenada, tranquila y disciplinada Entorno físico atractivo, limpio, bien cuidado y agradable. 	<ul style="list-style-type: none"> Orientación del clima escolar hacia la eficacia y las relaciones interpersonales Establecimiento de políticas de disciplina claras y consistentes entre profesores y directivos

Personal directivo	Liderazgo	
	<ul style="list-style-type: none"> • Cualidades de gestión y organización • Cualidades de planificación, participación e implicación con la comunidad educativa • Capacidad para la toma de decisiones participativas y coordinadas • Capacidad de control de calidad y manejo de grupos • Promoción del trabajo en equipo • Corresponsabilidad de funciones y metas con el equipo directivo y el grupo docente • Participación de los docentes en la toma de decisiones. • Promoción de incentivos, reconocimiento y otros elementos de motivación 	<ul style="list-style-type: none"> • Comunicación de las altas expectativas a los alumnos y reconocimiento a rendimientos excelentes • Apoyo a la profesión docente y a la operación de los procesos de la escuela • Promoción y planteamiento de objetivos consensuados • Firmeza y capacidad de establecer propósitos claros • Promoción de la unidad de propósito, del trabajo colegiado y la colaboración, para lograr dicha unidad de propósitos • Compromiso de directivos y docentes en actividades de desarrollo profesional y de aprendizaje colectivo
	Expectativas altas	
	<ul style="list-style-type: none"> • Enfoque hacia el rendimiento en las materias básicas • Comunicar las expectativas • Altas expectativas con respecto del docente y del alumno 	<ul style="list-style-type: none"> • Énfasis en lo académico y centrado en el rendimiento • Tener altas expectativas globales, lo que implica una directiva adecuada a esas altas expectativas
	Gestión del tiempo	
	<ul style="list-style-type: none"> • Planeación de actividades del centro 	<ul style="list-style-type: none"> • Promoción de administración del tiempo por parte de los docentes
	Atención a la diversidad	
	<ul style="list-style-type: none"> • Adaptación a estudiantes con necesidades especiales • Formación y sensibilidad multicultural • Identificación de alumnos en riesgo de abandono escolar; atención y desarrollo de actividades para su permanencia 	<ul style="list-style-type: none"> • Utilización de prácticas validadas para la prevención del consumo de tabaco, alcohol y drogas • Colaboran con las agencias comunitarias para apoyar a las familias con necesidades sanitarias o sociales urgentes
Factores psicopedagógicos	Currículo de alta exigencia adaptada	
	<ul style="list-style-type: none"> • Enfoque claro del currículo • Planeación apropiada al currículo • Inclusión de la preparación para el trabajo entre las metas escolares • Énfasis en el desarrollo de habilidades básicas y de orden superior • Promoción del respeto y la empatía entre alumnos con diferentes antecedentes socio-económicos y culturales 	<ul style="list-style-type: none"> • Oferta de tiempo e instrucción extra a los alumnos con altas necesidades • Apoyo a la capacidad de adaptación (social y académicamente) de los alumnos con altas necesidades • Opciones y metodología de oportunidad para aprender

	Planeación educativa	
	<ul style="list-style-type: none"> Relaciones entre estudiantes y estudiante-docente Disciplina y control en el aula 	<ul style="list-style-type: none"> Capacidades de trabajo en el aula y trabajo en equipo Satisfacción de alumnos y maestros
	Uso de la evaluación y la retroalimentación	
	<ul style="list-style-type: none"> Frecuente y sistemático seguimiento de los resultados de alumnos y grupos para: (1) saber si los objetivos del centro se están alcanzando, (2) centrar la atención de los docentes, alumnos y padres en dichos objetivos, (3) optimizar la planificación, los métodos didácticos y la evaluación y (4) hacer evidente el interés de la institución en el progreso de los estudiantes 	<ul style="list-style-type: none"> Uso de los resultados de la evaluación Evaluación sistemática de la actuación del centro, entrega de resultados y su aprovechamiento en la toma de decisiones Uso de tecnología de evaluación Promoción del uso de evaluaciones alternativas, además de las pruebas tradicionales Estrategias para evitar el rezago de los estudiantes

Escuela (Aula)		
Factores del aula	Infraestructura del aula	
	<ul style="list-style-type: none"> Dotación y calidad del aula 	<ul style="list-style-type: none"> Recursos curriculares
	Clima en el aula	
	<ul style="list-style-type: none"> Relaciones entre estudiantes y estudiante-docente Disciplina y control en el aula 	<ul style="list-style-type: none"> Capacidades de trabajo en el aula y trabajo en equipo Satisfacción de alumnos y maestros
Factores del docente	Características del docente	
	<ul style="list-style-type: none"> Cualidades docentes Formación inicial y formación continua o mejora permanente Experiencia docente Planificación docente para el trabajo en aula Metodología y estructuración de las clases Preparación de las clases Recursos materiales y didácticos disponibles Utilización de materiales y recursos didácticos Uso de tecnología informática para el apoyo instructivo y como forma de acercarse al trabajo productivo 	<ul style="list-style-type: none"> Tutoría y seguimiento Empatía, implicación o relación e interacción personal de los profesores hacia los estudiantes Aporte de incentivos, reconocimiento y recompensas para promover la excelencia Altas expectativas del profesor sobre el grupo de alumnos Satisfacción del profesor Número de alumnos por profesor (ratio maestro-alumno) Estabilidad en el trabajo y condiciones laborales Responsabilidades y metas compartidas con el equipo directivo y con el resto del grupo docente

	Expectativas altas	
	<ul style="list-style-type: none"> • Enfoque hacia el rendimiento en las materias básicas • Altas expectativas del docente • Registro del avance de los estudiantes • Altas expectativas globales y enseñanza adecuada a ellas 	<ul style="list-style-type: none"> • Altas expectativas educativas con relación al alumno • Comunicación de las expectativas. • Énfasis en lo académico • Enseñanza centrada en el rendimiento • Aportar desafíos intelectuales a todos los alumnos en todas las clases
	Atención a la diversidad	
	<ul style="list-style-type: none"> • Adaptación a estudiantes con necesidades especiales • Enseñanza adaptativa 	<ul style="list-style-type: none"> • Atención a alumnos con riesgo • Formación y sensibilidad multicultural
	Gestión del tiempo	
	<ul style="list-style-type: none"> • Planeación de actividades • Manejo apropiado del tiempo de docencia 	<ul style="list-style-type: none"> • Administración del tiempo efectivo para favorecer el aprendizaje
	Uso de la evaluación y la retroalimentación	
	<ul style="list-style-type: none"> • Cualidades del docente para evaluación y seguimiento • Evaluación continua, incluyendo diagnóstico, retroalimentación y seguimiento de los resultados de los alumnos • Supervisión del progreso de los alumnos de forma cercana 	<ul style="list-style-type: none"> • Uso de tecnología de evaluación • Desarrollo y uso de evaluaciones alternativas, además de las pruebas tradicionales • Evaluación de aprendizajes de orden superior • Uso de los resultados de la evaluación
	Aprendizaje	
	<ul style="list-style-type: none"> • Énfasis en el aprendizaje de estrategias de orden superior 	<ul style="list-style-type: none"> • Promoción de técnicas de aprendizaje independiente y en equipo • Retroalimentación y refuerzo positivo

Estudiante		
Factores de la familia	Características de la familia y el hogar	
	<ul style="list-style-type: none"> • Estructura familiar • Nivel socioeconómico familiar • Nivel cultural de los padres • Satisfacción de los padres • Infraestructura material y físicas de la casa • Distancia hogar-escuela • Recursos educativos en el hogar 	<ul style="list-style-type: none"> • Implicación de la familia o interacción para el aprendizaje del estudiante • Participación de los padres en asociaciones • Relación de los padres con la escuela • Contacto de los directivos con los padres

Factores del estudiante	Características del estudiante	
	<ul style="list-style-type: none">• Sexo o género• Edad, nativo/inmigrante, lengua materna• Asistencia a educación inicial• Autoconcepto• Comportamiento	<ul style="list-style-type: none">• Convivencia social• Satisfacción con la escuela• Hábitos culturales• Aspiraciones académicas
	Gestión del tiempo	
	<ul style="list-style-type: none">• Administración del tiempo efectivo para el aprendizaje	<ul style="list-style-type: none">• Planeación de actividades
	Expectativas altas	
	<ul style="list-style-type: none">• Expectativas educativas del alumno• Énfasis en lo académico	<ul style="list-style-type: none">• Aceptar desafíos intelectuales en las clases
	Aprendizaje	
	<ul style="list-style-type: none">• Asistencia a la escuela• Tareas escolares realizadas en casa• Realización de actividades extraescolares• Hábitos de responsabilidad en el aprendizaje dentro de la escuela	<ul style="list-style-type: none">• Esperanza de vida académica según el profesor• Preocupación por elevar su autoestima• Interacción con pares de mayor rendimiento• Trabajo controlado de los alumnos

Anexo 2. Especificaciones para los modelos del estudio

A2.1 Variables

Como ya se indicó al describir el proyecto PISA 2006, los datos correspondientes a las variables descriptivas o explicativas se recogieron por medio de cuestionarios de datos contextuales (OECD-PISA, 2007 y 2008), asociados por una parte con datos de los estudiantes (familiares, sociodemográficos, culturales) y por otra parte con datos de la escuela (organización, datos curriculares). Es evidente que el modelo de calidad está contenido de manera explícita o implícita en el proyecto PISA 2006 pero no necesariamente organiza los posibles constructos de acuerdo con el modelo del INEE.

Si se acepta que hay un modelo CIPP subyacente al proyecto PISA, las variables recogidas con los cuestionarios no solamente se refieren a aspectos contextuales, ya que también tratan de obtener información sobre los procesos que pueden estar asociados con alguna de las dimensiones de la calidad educativa.

Las preguntas, cuestiones, rasgos o aspectos contenidos en los cuestionarios del proyecto PISA se pueden dividir en dos tipos básicos:

- A) Variables que proporcionan datos crudos (respuestas a preguntas específicas)
- B) Variables que se procesan para producir indicadores (combina las respuestas de varias preguntas, establece una métrica principalmente con el modelo de Rasch y produce un índice).



Dentro de las peculiaridades que se tienen en la recolección de datos en los cuestionarios, también se pueden distinguir estos casos:

- C) Variables que cuentan con la opción para que cada país establezca categorías o sub-variables propias, con el propósito de regionalizar la información de manera conveniente al contexto nacional.
- D) Variables categóricas que se transforman en una bandera (variable muda o dummy). Una bandera es una variable numérica, que transforma las categorías de una variable cualitativa para permitir hacer cálculos con los modelos de regresión y los modelos multinivel. Tal es el caso de la variable sexo o género, codificada como 0 varón y 1 mujer. Al emplear este tipo de variables, es posible obtener medidas estadísticas diversas, que no necesariamente se asimilan a una medida dentro de las categorías, sino solamente como proporción, como puede ser el caso de una media de género de 0.58, que indicaría la proporción de mujeres en la muestra, es decir, 58% de personas son mujeres, de acuerdo con la codificación indicada.
- E) Variables tipificadas o estandarizadas, que hacen una descripción de valores en términos de desviaciones estándar respecto de la media.
- F) Variables centradas, que solo hacen una traslación de la media hacia el valor 0, pero cuyos valores pueden o no hacer referencia a desviaciones estándar.

Cada una de las variables aparecen dentro de los cuestionarios una vez operacionalizados los constructos, de acuerdo con el manual de diseño del proyecto.

Para ser congruentes con las recomendaciones del proyecto PISA, se dio preferencia a las preguntas que conducen a indicadores, los cuales están incluidos dentro de la misma base de datos y, como se indica en los materiales de referencia del proyecto, no tienen que ser reprocesadas o adaptadas para los estudios que se efectúen en cada país, salvo los casos en que hubiera omisiones, que deben ser tratados por procesos de imputación como se indica posteriormente.

A2.1.1 Descripción de las variables disponibles en las bases de datos de PISA y otros datos tomados de estudios propios del INEE

Tabla A2.1 Variables nivel estudiante

Variable	Etiqueta	Descripción	Variable de imputación	Porcentaje de imputación
NOM_EDO	Nombre del estado	Nombre del estado de la República, en mayúsculas	N.A.	N.A.
REGION	Región	Identificador de la región	N.A.	N.A.
SCHOOLID	School ID 5-digit	Identificador de la escuela	N.A.	N.A.
ESCID	Identificador de la escuela por región	Se forma a partir de la región y el número de la escuela: $EscID = 100000 \times Region + SCHOOLID$	N.A.	N.A.
STIDSTD	Student ID 5-digit	Identificador del estudiante	N.A.	N.A.
ESTUID	Identificador del estudiante por región escuela	Se forma a partir de ESCID y del número del estudiante $EstuID = 1000 \times EscID + STIDSTD$	N.A.	N.A.
NIVEL	Nivel	Nivel de estudios del estudiante: 0=Secundaria, 1=Educación media superior	N.A.	N.A.
SC02Q01	Public or private Q2	Tipo de sostenimiento de la escuela: 0=Público, 1=Privado	N.A.	N.A.
URBANORURAL	Comunidad urbano rural	Tipo de comunidad donde se encuentra la escuela: 0=Rural, 1=urbano	N.A.	N.A.
W_FSTUWT	FINAL STATE-BASED STUDENT WEIGHT	Ponderador del estudiante. Se utiliza para expandir los resultados hacia la población	N.A.	N.A.
PVSCIE	Plausible value in science	Primer valor plausible del puntaje obtenido en la prueba, de acuerdo con la distribución <i>a posteriori</i> de resultados de PISA 2006	N.A.	N.A.
PV2SCIE	Plausible value in science	Segundo valor plausible del puntaje obtenido en la prueba, de acuerdo con la distribución <i>a posteriori</i> de resultados de PISA 2006	N.A.	N.A.
PV3SCIE	Plausible value in science	Tercer valor plausible del puntaje obtenido en la prueba, de acuerdo con la distribución <i>a posteriori</i> de resultados de PISA 2006	N.A.	N.A.
PV4SCIE	Plausible value in science	Cuarto valor plausible del puntaje obtenido en la prueba, de acuerdo con la distribución <i>a posteriori</i> de resultados de PISA 2006	N.A.	N.A.
PV5SCIE	Plausible value in science	Quinto valor plausible del puntaje obtenido en la prueba, de acuerdo con la distribución <i>a posteriori</i> de resultados de PISA 2006	N.A.	N.A.
GENERO	Género	Género del estudiante. Codificación: 0=hombre, 1=mujer.	N.A.	N.A.
REZAGO	Rezago	Se considera que un estudiante es rezagado si cumplió años en el periodo de enero-agosto de 1990 y está inscrito en secundaria. Codificación: 0=Sin rezago, 1=Rezagado	N.A.	N.A.

Variable	Etiqueta	Descripción	Variable de imputación	Porcentaje de imputación
ESCS	Index of economic, social and cultural status PISA 2006 (WLE)	Es un índice que se construye para cada estudiante, teniendo en cuenta cinco índices, estos son: - Nivel educativo alcanzado por alguno de los padres en años de estudio (PARED) - Ocupación de mayor categoría entre los padres (HISEI) - Bienes materiales con los que cuenta la familia del estudiante (en este índice se incluyen las tres preguntas realizadas por cada país sobre bienes en el hogar) (WEALTH) - Posesiones con valor cultural en el hogar (CULTPOSS) - Recursos educativos en el hogar (HEDRES) En el informe <i>Questionnaire indices for PISA 2006</i> puede consultarse más detalles sobre la construcción del índice	BaEscs	0.3%
ESCS2	ESCS2	Es el cuadrado del ESCS que se emplea en varios modelos HLM para reportar la influencia de la no linealidad del ESCS en el rendimiento de la prueba	N.A.	N.A.
PESCS	Percentil ESCS	Se creó una variable discreta del 1 al 10 con los deciles del ESCS. Después se asignó a cada estudiante el número de decil que le corresponde de acuerdo con su ESCS	N.A.	N.A.
HOMEPOS	Index of home possessions PISA 2006 (WLE)	Índice de posesiones en el hogar compuesto por 3 índices 1) Bienestar familiar (WEALTH) 2) Posesiones con valor cultural en el hogar (CULTPOSS) 3) Recursos educativos en el hogar (HEDRES)	BaHomePos	0.2%
ENVOPT	Environmental optimism PISA 2006 (WLE)	Optimismo respecto al medio ambiente ST25Q01 Contaminación atmosférica ST25Q02 Escasez de energía ST25Q03 Extinción de plantas y de animales ST25Q04 Tala de los bosques para la otra utilización del suelo ST25Q05 Encasque de agua ST25Q06 Basura nuclear	BaEnvOpt	0.5%
HEDRES	Home educational resources PISA 2006 (WLE)	Recursos educativos en el hogar ST13Q01 Un escritorio o mesa para estudiar ST13Q03 Un lugar tranquilo para estudiar ST13Q04 Una computadora que puedas usar para tus tareas escolares ST13Q05 Programas educativos para la computadora ST13Q07 Tu propia calculadora ST13Q11 Libros de consulta para tus tareas escolares ST13Q12 Un diccionario	BaHedres	0.7%
WEALTH	Family wealth PISA 2006 (WLE)	Índice de bienestar familiar ST13Q02 Una habitación sólo para ti ST13Q06 Una conexión a Internet ST13Q13 Una lavadora de platos ST13Q14 Un DVD o videocasetera ST14Q01 Número de teléfonos celulares que poseen en casa ST14Q02 Número de televisores que poseen en casa ST14Q03 Número de computadoras que poseen en casa ST14Q04 Número de autos que posee en casa Además de otros tres artículos específicos según el país, para México fueron: ST13Q15 Servicio de televisión de paga (Sky, Cablevisión, etc.) ST13Q16 Línea telefónica ST13Q17 Horno de microondas	BaWealth	0.3%

Variable	Etiqueta	Descripción	Variable de imputación	Porcentaje de imputación
ENVAWARE	Awareness of environmental issues PISA 2006 (WLE)	Conocimiento de aspectos medio ambientales ST22Q01 Aumento de los gases de efecto invernadero en la atmósfera ST22Q02 Uso de organismos genéticamente modificados ST22Q03 Lluvia ácida ST22Q04 Basura nuclear ST22Q05 Las consecuencias de la tala de los bosques y explotación de suelos	BaEnVaWare	0.4%
RESPDEV	Responsibility for sustainable development PISA 2006 (WLE)	Conciencia del estudiante para el desarrollo sustentable ST26Q01 Es importante hacer verificaciones periódicas de las emisiones de gases de los coches como una condición para su uso ST26Q02 Me disgusta que se desperdicie la energía eléctrica cuando se ponen a funcionar aparatos sin motivo alguno ST26Q03 Estoy en favor de que haya leyes que regulen los desperdicios de las fábricas, aun cuando esto incremente el precio de los productos ST26Q04 Debería reducirse al mínimo el uso de envases de plástico para disminuir la cantidad de basura ST26Q05 Se debería obligar a las fábricas a demostrar que eliminan sus desechos peligrosos en condiciones totalmente seguras. ST26Q06 Estoy en favor de que haya leyes que protejan el hábitat de las especies en peligro de extinción ST26Q07 Siempre que sea posible se debe producir electricidad a partir de fuentes renovables, aun cuando esto aumente el costo	BaRespDev	0.7%
SCIEEFF	Science self-efficacy PISA 2006 (WLE)	Auto eficacia ST17Q01 Reconocer la parte científica que hay en un artículo sobre salud publicado en un periódico ST17Q02 Explicar por qué los terremotos se dan en algunas zonas con más frecuencia que en otras. ST17Q03 Describir la función de los antibióticos en el tratamiento de una enfermedad ST17Q04 Determinar el tema científico relacionado con el tratamiento de la basura ST17Q05 Predecir cómo los cambios ambientales podrían afectar la supervivencia de ciertas especies ST17Q06 Interpretar la información científica que viene en las etiquetas de artículos comestibles ST17Q07 Discutir por qué nuevas evidencias pueden llevar a cambiar tu opinión respecto a la posibilidad de que hubiera vida en Marte ST17Q08 Entre dos explicaciones acerca de la formación de lluvia ácida, identificar cuál es la mejor	BaScieEff	0.3%
ENVPERC	Perception of environmental issues PISA 2006 (WLE)	Percepción del estudiante sobre temas ambientales ST24Q01 ST24Q02 Contaminación del aire ST24Q03 Escasez de energéticos ST24Q04 Extinción de plantas y animales ST24Q05 Tala de bosques para darle otro uso a la tierra ST24Q06 Escasez de agua ST24Q06 Desechos nucleares	BaEnvPerc	0.6%

Variable	Etiqueta	Descripción	Variable de imputación	Porcentaje de imputación
CULTPOSS	Cultural possessions at home PISA 2006 (WLE)	Posesiones culturales en el hogar ST13Q09 Libros de poesía ST13Q10 Obras de arte (ejemplo=, pinturas) ST13Q08 Libros de literatura clásica (ejemplo= El Quijote, Cervantes;)	BaCultPoss	2.5%
GENSCIE	General value of science PISA 2006 (WLE)	El valor que tiene la ciencia para los estudiantes ST18Q01 Los avances en ciencia y tecnología normalmente mejoran las condiciones de vida de las personas ST18Q02 La ciencia es importante para ayudarnos a entender el mundo natural que nos rodea ST18Q04 Los avances en ciencia y tecnología normalmente ayudan a mejorar la economía ST18Q06 La ciencia es valiosa para la sociedad ST18Q09 Los avances en ciencia y tecnología normalmente traen beneficios sociales	BaGenScie	0.5%
SCINVEST	Science Teaching - Student investigations PISA 2006 (WLE)	Enseñanza de la ciencia basada en trabajos de investigaciones ST34Q08 Se permite a los estudiantes diseñar sus propios experimentos ST34Q11 El maestro da oportunidad a los estudiantes de escoger sus propias investigaciones ST34Q16 Los estudiantes realizan una investigación para demostrar sus propias ideas	BaScInvest	2.7%
HISEI	Highest parental occupational status	Ocupación de mayor jerarquía entre los padres del estudiante	BaHiSei	3.3%
HISCED	Highest educational level of parents	Grado académico de mayor jerarquía entre los padres del estudiante	BaHiScEd	0.4%
PARED	Highest parental education in years	Grado académico de mayor jerarquía entre los padres del estudiante medido en años escolares	BaParEd	0.4%
CARINFO	Student information on science-related careers PISA 2006 (WLE)	Información sobre carreras relacionadas con las Ciencias ST28Q01 Carreras relacionadas con la ciencia que existen en el mercado laboral ST28Q02 Dónde encontrar información sobre carreras relacionadas con la ciencia ST28Q03 Los pasos que deben darse para seguir una carrera relacionada con la ciencia ST28Q04 Empleadores o compañías que contratan gente con carreras relacionadas con la ciencia	BaCarInfo	0.7%
CARPREP	School preparation for science-related careers PISA 2006 (WLE)	Formación que da la escuela para las carreras relacionadas con las Ciencias ST27Q01 Las materias que se imparten en mi escuela permiten que los estudiantes tengan habilidades y conocimientos necesarios para seguir una carrera relacionada con la ciencia ST27Q02 Las clases de Ciencias Naturales que se imparten en mi escuela permiten que los estudiantes tengan habilidades y conocimientos básicos para elegir diferentes carreras ST27Q03 Las materias que estudio me dan habilidades y conocimientos básicos para seguir una carrera relacionada con la ciencia ST27Q04 Mis maestros me dan las aptitudes básicas y los conocimientos necesarios para seguir una carrera relacionada con la ciencia	BaCarPrep	0.7%

Variable	Etiqueta	Descripción	Variable de imputación	Porcentaje de imputación
JOYSCIE	Enjoyment of science PISA 2006 (WLE)	Disfruto aprender temas relacionadas con las ciencias ST16Q01 Me resulta agradable aprender cosas de ciencias en general ST16Q02 Disfruto cuando leo sobre ciencias en general ST16Q03 Soy feliz resolviendo problemas de ciencias en general ST16Q04 Me gusta saber cosas nuevas sobre ciencias en general ST16Q05 Tengo interés en aprender sobre ciencias en general	BaJoyScie	0.3%
PERSCIE	Personal value of science PISA 2006 (WLE)	Valor personal de los estudiantes hacia las ciencias ST18Q03 Algunos conceptos de ciencia me ayudan a ver cómo me relaciono con otras personas ST18Q05 Usaré la ciencia de muchas maneras cuando sea adulto ST18Q07 La ciencia es muy importante para sí ST18Q08 Creo que la ciencia me ayuda a entender las cosas que me rodean ST18Q10 Cuando termine la escuela tendré muchas oportunidades de usar la ciencia	BaPerScie	0.6%
SCAPPLY	Science Teaching - Focus on applications or models PISA 2006 (WLE)	Enseñanza de las ciencias a través de aplicaciones o modelos ST34Q07 El maestro explica cómo una fórmula o idea científica puede aplicarse a distintos fenómenos (ejemplo= el movimiento de los objetos, sustancias con propiedades similares) ST34Q12 El maestro se apoya en las Ciencias Naturales para ayudar a los estudiantes a entender el mundo que los rodea ST34Q15 El maestro explica claramente la importancia del concepto de ciencia en general en nuestras vidas ST34Q17 El maestro usa ejemplos de aplicación tecnológica para mostrar por qué la ciencia en general es importante para la sociedad	BaScApply	2.7%
SCHANDS	Science Teaching - Hands-on activities PISA 2006 (WLE)	Enseñanza de la ciencia con actividades manuales ST34Q02 Los estudiantes tienen tiempo para realizar experimentos prácticos en el laboratorio ST34Q03 Se pide a los estudiantes que propongan cómo se podría investigar en el laboratorio un tema de ciencias ST34Q06 Se pide a los estudiantes que obtengan conclusiones sobre un experimento que hayan realizado ST34Q14 Los estudiantes hacen experimentos siguiendo las instrucciones del maestro	BaScHands	2.5%
SCIEFUT	Future-oriented science motivation PISA 2006 (WLE)	Motivación orientada hacia las ciencias ST29Q01 Quisiera trabajar en una profesión que tuviera que ver con la ciencia en general ST29Q02 Después de terminar esta escuela, quisiera estudiar ciencia en general ST29Q03 Me gustaría pasarme la vida trabajando en materias de ciencia avanzada ST29Q04 Me gustaría trabajar en proyectos de ciencia en general cuando sea adulto	BaScieFut	0.7%

Variable	Etiqueta	Descripción	Variable de imputación	Porcentaje de imputación
SCSCIE	Science self-concept PISA 2006 (WLE)	Autoconcepto en ciencias ST37Q01 El aprendizaje avanzado de Ciencias Naturales sería fácil para mí ST37Q02 Generalmente, en los exámenes, contesto bien las preguntas de Ciencias Naturales ST37Q03 Aprendo rápidamente temas de Ciencias Naturales ST37Q04 Los temas de Ciencias Naturales son fáciles para mí ST37Q05 Cuando me enseñan Ciencias Naturales, entiendo muy bien los conceptos ST37Q06 Entiendo fácilmente las nuevas ideas sobre Ciencias Naturales	BaScScie	4.0%

Nota: N.A. se refiere a que no aplica realizar el cálculo, en virtud de que la variable no presentó valores perdidos.

Tabla A2.2 Variables nivel escuela

Variable	Etiqueta	Descripción	Variable de imputación	Porcentaje de imputación
NOM_EDO		Nombre del estado de la República, en mayúsculas	N.A.	N.A.
REGION	Región	Identificador de la región	N.A.	N.A.
SCHOOLID	School ID 5-digit	Identificador de la escuela	N.A.	N.A.
ESCID		Se forma a partir de la región y el número de la escuela:	N.A.	N.A.
SCIPROM	School activities to promote the learning of science (WLE)	Actividades de la escuela para promover el aprendizaje de las ciencias entre los estudiantes SC20Q01 Clubes de ciencias SC20Q02 Ferias de ciencias SC20Q03 Concursos de ciencias SC20Q04 Proyectos de ciencias extracurriculares (incluye proyectos de investigación) SC20Q05 Excursiones y trabajo de campo	BaSciProm	3.5%
SCMATEDU	Quality of educational resources (WLE) (2006)	Calidad de los recursos educativos de la escuela SCI4Q07 Equipo para laboratorio de ciencias escaso o inadecuado SCI4Q08 Material de enseñanza escaso o inadecuado (ejemplo= libros de texto) SCI4Q09 Computadoras para enseñanza escasas o inadecuadas SCI4Q10 Conexiones a Internet escasas o inadecuadas 1 2 3 4 SCI4Q11 Programas de computadora para la enseñanza escasos o inadecuados SCI4Q12 Materiales de biblioteca escasos o inadecuados SCI4Q13 Recursos de material audiovisual escasos o inadecuados	BaScMatEdu	1.7%
CLSIZE	Size of <test language> class recoded from SC06Q01	Promedio del tamaño de clase SC06Q01 15 o menos alumnos 16-20 alumnos 21-25 alumnos 26-30 alumnos 31-35 alumnos 36-40 alumnos 41-45 alumnos 46-50 alumnos Más de 50 alumnos	BaClrSize	6.9%
ENVLEARN	School activities for learning environmental topics (WLE)	Actividades en la escuela para aprender temas sobre el medio ambiente a) Prácticas de campo 1 2 b) Visitas a museos 1 2 c) Visitas a centros científicos y tecnológicos 1 2 d) Proyectos del medio ambiente extracurriculares (incluye proyectos de investigación) 1 2 e) Conferencias y/o seminarios (por ejemplo, los impartidos por conferencistas externos) 1 2	BaEnvLear	3.0%
PROPCERT	Proportion of certified teachers	Proporción de los profesores con licenciatura. Es el número de los profesores completamente certificados	BaPropCert	49.1%
XESCS	Media del ESCS a nivel escuela	Se asignó a cada escuela el promedio del ESCS de los estudiantes que estaban inscritos en la misma	N.A.	N.A.
NIVEL	Nivel	Nivel de estudios que imparte la escuela 0= Secundaria 1= Educación media superior	N.A.	N.A.
SC02Q01	Public or private Q2	Tipo de sostenimiento de la escuela: 0= Público, 1=Privado	N.A.	N.A.
URBANORURAL	Comunidad urbano rural	Tipo de comunidad en la que se encuentra la escuela: 0=Rural, 1=Urbano	N.A.	N.A.

Tabla A2.3 Variables de la base de datos de regiones (nivel estado-región)

VARIABLE	Etiqueta y categorías	Descripción	Rango de valores	Etiqueta de imputación y porcentaje de imputaciones
REGION	Región	Las 32 entidades de la República Mexicana se agruparon en siete regiones de acuerdo con sus características geográficas y socioeconómicas. Se considera que estas regiones son unidades con desarrollo económico y social semejantes por lo que se espera tener una comprensión más robusta de los resultados de las evaluaciones educativas realizadas en ellas. Codificación: 1= Región Sur-Oeste, 2= Región Distrito Federal, 3= Región Sur-Este, 4= Región Norte, 5= Región Centro-Norte, 6= Región Occidente, 7= Región Centro-Sur	1 - 7	N.A.
PIB PER CAPITA	PIB <i>per cápita</i>	Producto interno bruto <i>per cápita</i> , en dólares PPA, en 2004 [PISA2006, Tabla 2.4, p. 27]	3,827 - 23,865	N.A.
IDH	Índice de Desarrollo Humano México 2006-2007	Es una medida resumen que promedia los avances relativos de un país o demarcación geográfica, sobre tres aspectos básicos del desarrollo humano requeridos por la población: gozar de una vida larga, contar con educación y disfrutar de un nivel de vida decoroso. [AM, Tabla 40, p. 90] [PISA2006, Tabla 2.4, p.27]	0.7 - 0.9	N.A.
EMIGRANTE	Porcentaje de emigrantes	Porcentaje que representa la población de emigrantes mexicanos que residen en Estados Unidos respecto a la población residente en México por región de nacimiento, 2005. [Migración México-Estados Unidos, Panorama regional y estatal. México]	0.7- 35.9	N.A.
GRESCOLARIDAD		Escolaridad en grados de toda la población [AM, Tabla 46, p. 62]	6.1 - 10.2	N.A.
ANALFAB	Analfabetismo	Porcentaje de personas analfabetas [AM, Tabla 46, p. 62]	1.4 - 19.4	N.A.
BIBLIOT			11.8 - 105.9	N.A.
GRSOBREANALF	Grado sobre índice de analfabetismo	Cociente que relaciona al grado de escolaridad con el analfabetismo	0.3 - 6.4	N.A.
GRXANALF	Grado por índice de analfabetismo	Producto del grado de escolaridad con el analfabetismo	12.5 - 123.8	N.A.
MATRICULA	Matrícula	Estudiantes de educación básica en las entidades federativas (INEE Panorama Educativo de México, 2005)	112,650 - 3,196,587	N.A.
GASTOFAEB	Gasto FAEB	Gasto FAEB en 2005 en millones de pesos a precios corrientes (RSI 2-I.1, p. 282)	1,646 - 19,416	N.A.
GASTOXALUMNO	Gasto por alumno	Gasto FAEB entre la matrícula en pesos a precios corrientes	5,013.2 - 15,194.0	N.A.
MARGINALIDAD	Índice de marginación 2005	Índice de marginación en 2005 (CS10, p.115) Es una medida que agrupa los valores relativos de nueve indicadores socioeconómicos, con la finalidad de diferenciar unidades territoriales según las carencias padecidas por la población como resultado de falta de acceso a la educación, residencia en viviendas inadecuadas, ingresos monetarios insuficientes y residencia en localidades pequeñas	-1.5 - 2.4	N.A.

DOCLICENC	Directores con licenciatura	Porcentaje de directores con licenciatura o más (RS03-1.1, p. 219)	0 - 100	N.A.
DOCCAPAC	Directores docentes	Porcentaje de directores con formación docente (RS03-1.1, p. 219)	0 - 100	N.A.
DOCCARRERA	Directores con magisterio	Porcentaje de directores con carrera magisterial (RS03-1.1, p. 219)	0 - 100	N.A.
DOCEXPER	Experiencia del director	Años de experiencia laboral de los directores (RS03-1.1, p. 219)	3.3 - 7.6	N.A.
COMPUPUBL	Porcentaje de públicas con computadoras	Porcentaje de escuelas públicas sin computadora (RS06a-1.3, p. 246)	7.2 - 65.8	N.A.
COMPUPRIV	Porcentaje de privadas con computadora	Porcentaje de escuelas privadas sin computadora (RS06a-1.3, p. 246)	0.4 - 2.8	N.A.
PTERMGRAL	Probabilidad de terminación secundaria	Probabilidad de que un alumno termine la secundaria general en edad reglamentaria (AT01-2, p. 306)	0.0 - 1.0	N.A.
PTERMTELE	Probabilidad de terminación telesecundaria	Probabilidad de que un alumno termine la telesecundaria en edad reglamentaria (AT01-2, p. 306)	0.0 - 1.0	N.A.
PTERMTEC	Probabilidad de terminación técnica	Probabilidad de que un alumno termine la secundaria técnica en edad reglamentaria (AT01-2, p. 306)	0.0 - 1.0	N.A.
PCONTHOMBRE	Probabilidad hombres continúe estudiando	Probabilidad de que un alumno varón inicie secundaria y continúe media superior en edad reglamentaria (AT02-2, p. 313)	0.0 - 1.0	N.A.
PCONTMUJER	Probabilidad mujeres continúe estudiando	Probabilidad de que una alumna inicie secundaria y continúe media superior en edad reglamentaria (AT02-2, p. 313)	0.0 - 1.0	N.A.
PCONTTOTAL	Probabilidad de continuar estudiando	Probabilidad de que un alumno (hombre o mujer) inicie secundaria y continúe media superior en edad reglamentaria (AT02-2, p. 313)	0.0 - 1.0	N.A.
COBERTURA	Cobertura	Tasa neta de cobertura al 2005 (AT04-1, p. 325)	0 - 102.6	N.A.
EDADNORMAL	Edad normativa	Porcentaje de población que asiste a 3° de secundaria en edad reglamentaria (AT10-2, p. 366)	0 - 100	N.A.
APROBGRAL	Tasa aprobación secundaria	Tasa de aprobación en secundaria general en el ciclo 2004/2005 (AT12-4, p. 389)	0 - 100	N.A.
APROBTELE	Tasa aprobación telesecundaria	Tasa de aprobación en telesecundaria en el ciclo 2004/2005 (AT12-4, p. 389)	0 - 100	N.A.
APROBTEC	Tasa aprobación técnica	Tasa de aprobación en secundaria técnica en el ciclo 2004/2005 (AT12-4, p. 389)	0 - 100	N.A.
MATGRAL	Desempeño en Matemáticas en los EXCALE	Puntaje promedio general de los alumnos de 3° de secundaria en los Excale de Matemáticas (RE01-4, p. 455)	200 - 800	N.A.
MATPRIV	Desempeño en Matemáticas en secundaria privada	Puntaje promedio de los alumnos de 3° de secundaria privada en los Excale de Matemáticas (RE01-4, p. 455)	200 - 800	BaMatPriv
MATPUBL	Desempeño en Matemáticas en secundaria pública	Puntaje promedio de los alumnos de 3° de secundaria pública en los Excale de Matemáticas (RE01-4, p. 455)	200 - 800	-0.41
ESPGRAL	Desempeño en Español en los EXCALE	Puntaje promedio general de los alumnos de 3° de secundaria en los Excale de Español (RE01-4, p. 462)	200 - 800	N.A.

ESPPRIV	Desempeño en Español en la secundaria privada	Puntaje promedio de los alumnos de 3° de secundaria privada en los EXCALE de Matemáticas (RE01-4, p. 462)	200 - 800	BaEspPriv
ESPPUBL	Desempeño en la secundaria pública	Puntaje promedio de los alumnos de 3° de secundaria pública en los EXCALE de Matemáticas (RE01-4, p. 462)	200 - 800	-0.41

Nota: N.A. se refiere a que no aplica realizar el cálculo, en virtud de que la variable no presentó valores perdidos.

Las tablas A2.1 a A2.3 presentan el conjunto de variables empleadas para este estudio, en particular los datos de las tablas A2.1 y A2.2 provienen de la base de datos de PISA, en cambio los datos de la tabla A2.3 (base de datos regional) están tomados de cuatro referencias y se identifican por medio de paréntesis rectangulares o circulares, además de algunas siglas o referencias a tablas contenidas en las publicaciones:

1. Díaz, Flores y Martínez, (2007) PISA 2006. Por ejemplo: [PISA2006, Tabla 2.4, p. 27].
2. Robles y Martínez (2006) Por ejemplo: (RS12-I.1, p. 282)
3. Aguayo (2008) Por ejemplo: [AM, Tabla 46, p. 62]
4. Otras cantidades entre paréntesis sin siglas pueden ser derivadas de las tres referencias anteriores.

Los datos de la tabla A2.3 se tienen originalmente por entidad federativa (31 estados y el Distrito Federal), los resultados se ponderaron para calcular los índices regionales y finalmente se estandarizaron (media 0, desviación estándar 1) para todas las regiones, con la base así construida se hicieron los análisis multinivel.

A2.1.2 Comentarios sobre las variables empleadas en el estudio

Algunas de las variables de la base de datos de PISA 2006 recibieron una atención especial desde el inicio de este estudio, como es el caso del ESCS (Nivel económico, social y cultural), el rezago, la región, la marginación y las variables no-cognitivas. A continuación se describen las consideraciones hechas sobre las variables de interés.

A2.1.2.1 Índice de nivel económico, social y cultural

El valor del índice ESCS se expresa como desviaciones estándar de una distribución centrada en 0 para el estudiante promedio de la OCDE. El uso del ESCS proporcionado por PISA 2006 requirió de una revisión debido a la presencia de un conjunto adicional de preguntas o categorías nacionales, solicitadas por México a través del INEE. La integración de este índice se hace combinando otros tres: HOMEPOS (que contiene la posesión de bienes en el hogar, junto con los ítems de las variables WEALTH, CULTPOS y HEDRES y libros en el hogar), HISEI (mayor grado de escolaridad de uno de los padres), PARED (mayor número de años de escolaridad de los padres).

De acuerdo con el manual de PISA 2006, la variable de posesión de bienes en el hogar HOMEPOS incluye: un cuarto propio y un lugar para estudio del estudiante, con escritorio, calculadora de escritorio, libros de texto y diccionario; además, en la casa se espera que haya una computadora dotada de *software* y conexión a internet, disponibilidad de otros libros (de literatura, poesía) y algún objeto de arte o contar con un conjunto de bienes materiales (teléfono celular, televisores, automóviles, lavadora de trastes). Es claro que la disponibilidad de lavadora de trastes o lava-vajillas, puede indicar un nivel de bienestar en general para cualquier hogar, sin embargo se puede pensar que hay otros bienes mucho más cercanos al contexto nacional, por lo que México, a través del INEE, propuso tres bienes adicionales dentro de este inventario: servicio de televisión de paga (citando los nombres de las compañías que ofrecen este servicio), línea telefónica y horno de microondas. Para construir la escala de la variable HOMEPOS se incluyeron los bienes nacionales en dos etapas: la primera etapa establece que la suma de los ítems debe ser nula, incluyendo solamente el conjunto de ítems comunes a todos los países; la segunda etapa agrega los bienes considerados propios del país. Se hicieron imputaciones a los componentes omitidos por los estudiantes en función de una regresión sobre las otras variables considerando un componente de error aleatorio.

En principio se pensó de que cada país debía adaptar el ESCS en función de los elementos incluidos para HOMEPOS en el cuestionario de PISA, pero se encontró que en el Manual Técnico de PISA se estipula que las preguntas nacionales son tomadas en cuenta al momento de preparar el índice para agregarlo a la base de datos. Si se toma en cuenta que las categorías nacionales representan sólo una fracción del conjunto de categorías exploradas por el proyecto, alguien podría pensar que el ESCS se forma con preguntas que no son de interés para el país, pero no es el caso. La integración de un ESCS particular para cada país, con una ponderación diferente para las categorías nacionales, haría poco comparable el índice entre países, en cuyo caso se generaría la necesidad de comparar el índice socioeconómico de uso independiente para México con los correspondientes a otros países. Un análisis de correlación permite ver que el impacto que tiene cada categoría específica propuesta por el país es de poco efecto para describir un índice socioeconómico.

A2.1.2.2 Rezago académico del estudiante

La variable rezago del estudiante es interesante, porque se tuvo que derivar a partir de la información disponible en la base de datos en función de la edad de la persona y del año escolar que cursa, de acuerdo con la edad reglamentaria para ubicarse en el nivel esperado.

Se asignó una variable muda que identifica el caso de rezago, que se presenta cuando una persona se encuentra en Secundaria en el año 2006, habiendo nacido entre enero y agosto de 1990. Los demás casos no corresponden a rezago, porque se trata de estudiantes de Secundaria que entraron en su edad reglamentaria al haber nacido después del mes de agosto de 1990 o los estudiantes que se encuentran en nivel medio superior.

A2.1.2.3 Regionalización de México

La regionalización de la República Mexicana que propone el Instituto Nacional para la Evaluación de la Educación (INEE) se deriva de índices que miden directa o indirectamente el nivel de desarrollo educativo en las entidades federativas. Las regiones que se determinan tienen características demográficas, sociales, económicas y educativas más o menos homogéneas.

El objetivo de establecer esta regionalización es para mostrar las diferencias en los resultados de las evaluaciones de PISA por zonas, esto es, se trata de mostrar el comportamiento del sistema educativo para un área más amplia que un estado.

Para efectuar esta regionalización se consideró la división política de México; la subregionalización establecida por el Instituto Nacional de Estadística Geografía e Informática (INEGI); el índice de marginación, el índice de desarrollo humano, el tamaño de la población y el PIB *per cápita* de cada entidad; además del índice socioeconómico y cultural evaluado en PISA 2006 para cada entidad. Para agrupar a las entidades federativas se utilizó la técnica estadística multivariante de análisis de conglomerados (*cluster analysis*).

Las siete regiones en las que se dividió el país son:

- Región Sur-Oeste
- Región Distrito Federal
- Región Sur-Este
- Región Norte
- Región Centro-Norte
- Región Occidente
- Región Centro-Sur

Las entidades comprendidas en cada región se muestran en seguida.

Región Centro-Norte

En esta región se ubican los estados de: Aguascalientes, Guanajuato, Querétaro, Durango, San Luis Potosí y Zacatecas.

Región Centro-Sur

Formada por los estados de México, Morelos, Puebla y Tlaxcala.

Región Distrito Federal

Compuesta únicamente por el Distrito Federal

Región Norte

Agrupar a los estados de Baja California, Baja California Sur, Coahuila, Chihuahua, Nuevo León, Sonora y Tamaulipas.

Región Occidente

La componen los estados de Jalisco, Colima, Nayarit y Sinaloa.

Región Sur-Este

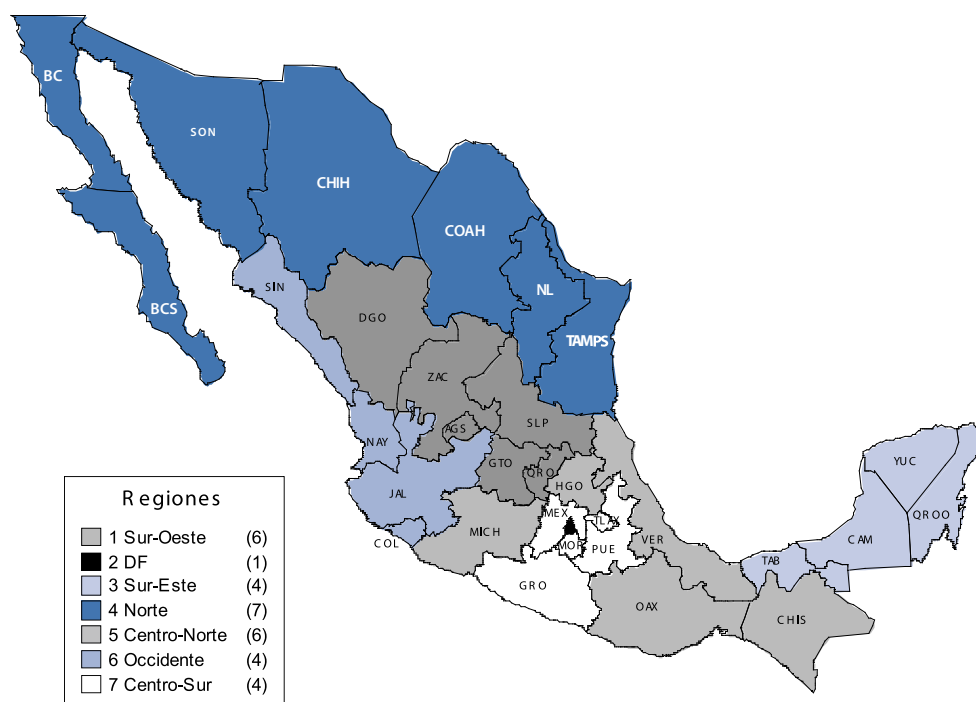
Está integrada por los estados de Campeche, Quintana Roo, Tabasco y Yucatán.

Región Sur-Oeste

Formada por los estados de Chiapas, Guerrero, Hidalgo, Michoacán, Oaxaca y Veracruz.

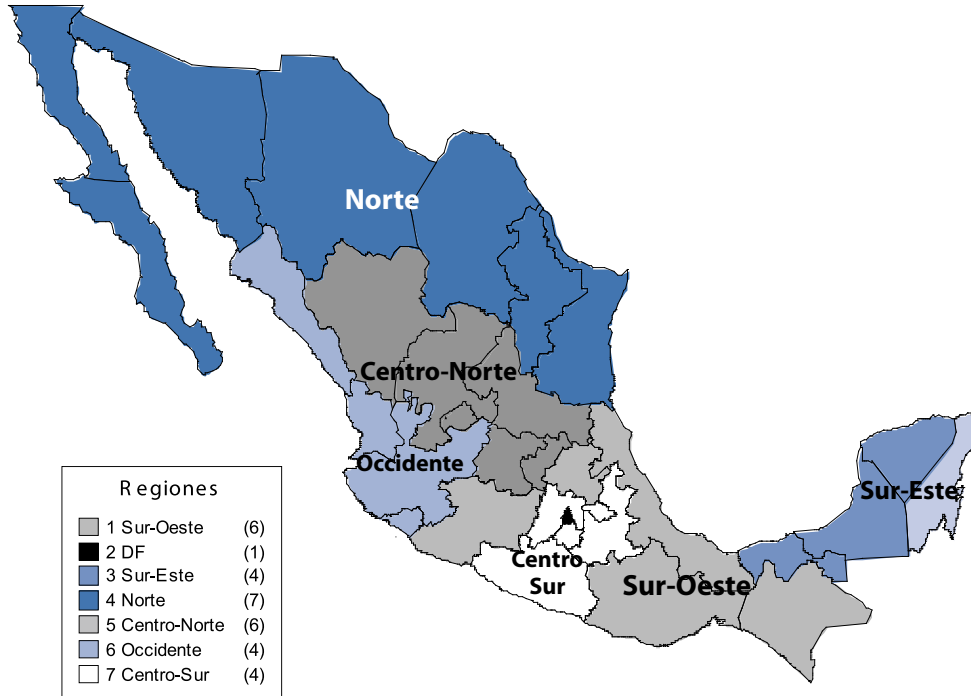
Estas regiones son entonces unidades con desarrollo educativo y social particulares en un espacio territorial, que ayudan a entender los resultados de las evaluaciones educativas. En cuanto a los nombres de las regiones ha de aclararse que, si bien corresponden a una división geográfica, estos no responden sólo a esta característica; los nombres sólo se asignaron para tener una referencia breve. En la figura A2.1 se muestran esquemáticamente los estados que quedan comprendidos dentro de cada región.

Figura A2.1 Regionalización socioeconómica propuesta por el INEE



La figura A2.2 es una propuesta alternativa para ilustrar la conformación de las regiones.

Figura A2.2 Regionalización socioeconómica propuesta por el INEE



A2.1.2.4 Marginación

Para el índice de marginación se pensaba utilizar el índice de GINI por estado, pero no se contó con el dato actualizado y producirlo queda fuera del propósito de este estudio que partió de utilizar solamente los datos disponibles en alguna fuente o que pudieran derivarse por agrupaciones de tales datos disponibles; se optó por emplear el índice de marginación disponible en el reporte del INEE denominado *Panorama Educativo de México* (Robles y Martínez, 2006).

A2.1.2.5 Ubicación decilar del estudiante

Para poder considerar la ubicación socioeconómica y cultural de cada estudiante en relación con el resto de escolares, se creó una variable discreta en función del decil en el cual se encuentra la persona respecto del ESCS. Ha sido costumbre que los estudios de PISA realicen comparaciones entre los estudiantes que se encuentran en 5% de las colas de la distribución, pero se sabe que para una distribución normal, la mayor diferencia entre subgrupos se presenta cuando se trabaja con las colas a 27% (Kelley, 1939). Para este estudio se tomó el acuerdo de considerar la ubicación decilar de los sujetos, atendiendo a que no todas las variables se distribuyen de forma normal y como un procedimiento sencillo de comparación entre grupos.












A2.1.2.6 Variables no cognitivas

Las variables no cognitivas, que representan actitudes para la ciencia o motivación al estudio, se tienen representadas parcialmente dentro del cuestionario del proyecto PISA 2006, por lo que se integran dentro de alguna de las dimensiones del modelo de calidad educativa. Se ha sugerido realizar un estudio para la predictividad de las variables no cognitivas respecto de los logros en las diferentes partes de la prueba, lo cual podría ser motivo de un trabajo posterior.

A2.1.3 Sobre el funcionamiento diferencial por género.

En el estudio PISA 2006 realizado por el INEE (Díaz, G.M.A., Flores V.G. y Martínez R.F., 2006, p.192 y 265 y sigs.) se presentan resultados del funcionamiento diferencial de los campos de Ciencias contenidos en la prueba, presentándose estos patrones diferenciales:

Tabla A2.4 Diferencias por género de variables no cognitivas de PISA

Área cognitiva	Diferencia significativa en favor de los hombres	Diferencia significativa en favor de las mujeres	Sin diferencia significativa por género
General de Ciencias			
Identificación de problemas científicos			
Explicación de fenómenos científicos			
Usos de evidencia científica			X
Área no cognitiva			
Conocimiento sobre la Ciencia			X
Tierra y Sistema espacial			
Sistemas vivientes			
Sistemas físicos			
Interés en aprender temas científicos			
Apoyo por investigación científica			X
Auto-eficacia en ciencia			
Auto concepto en Ciencias			X
Valoración general de la Ciencia			
Interés general en Ciencias			
Valoración personal de la Ciencia			

En el estudio multinivel se tomará la presencia del funcionamiento diferencial para considerarlo como una fuente de interés en los logros en Ciencias de los estudiantes.

A2.2 Imputación de datos faltantes

A2.2.1 Sobre las omisiones en la recopilación de datos del proyecto PISA 2006

Todo modelo experimental establece el conjunto de variables y el tamaño de muestra de acuerdo con una significancia y un poder estadístico, de tal forma que se logre la representatividad de la población en estudio. En toda base de datos procedente de un levantamiento muestral o poblacional, siempre se tiene una cantidad de omisiones por diversas causas: el ítem no es aplicable en todos los casos, el sujeto que debe responder ignora el aspecto solicitado, deficiencias en la aplicación o rechazo a responder de parte de algunas personas. Para evitar tener registros o campos faltantes, al diseñar la muestra se considera un porcentaje de sujetos adicionales, con lo que se aumenta el tamaño de muestra por mermas de diversos tipos, incluyendo atrición.

No es diferente el caso del proyecto PISA, por ello se tiene una cierta proporción de omisiones en algunos datos de los cuestionarios, bien sea porque los estudiantes y la familia desconocían lo que se solicitaba o simplemente porque evitaron responder. En el caso del cuestionario para las escuelas, las omisiones de datos que no proporciona el director tienen un impacto crucial para caracterizar a la institución.

Un análisis de omisiones en la base de datos muestra que:

- A) Hay pocos casos de datos omitidos por variable (pregunta o índice), lo cual favorece la realización de análisis multinivel con un pequeño impacto por faltantes (en la base de datos de los estudiantes 50% de las preguntas tienen menos de 1.5% de omisiones; contra 40% de preguntas con menos de 1.5% de omisiones en la base de datos de escuelas).
- B) Existe 15% de preguntas sin respuesta en la base de datos de escuela; estas variables corresponden con preguntas no aplicables al caso de México, pero incluidas en la base de datos de PISA 2006. Su presencia no interesa para los fines de este estudio.
- C) En complemento al *patrón de omisiones* se tiene un *patrón de reactivos* respondidos de la misma forma por todos los directivos escolares, lo cual es atribuible al contenido de los cuestionarios. La explicación de estos *patrones* es atribuible al esquema de gestión de las instituciones públicas, para las cuales el modelo de trabajo de países europeos no corresponde a México.

A2.2.2 Imputación en casos faltantes

Las omisiones en la base de datos de respuestas reduce las posibilidades de inferir tendencias no sesgadas sobre los datos, especialmente en los análisis jerárquicos multinivel. El problema de las omisiones fue analizado de manera sistemática por

Rubin (1976), produciendo varias metodologías de imputación de datos que evitan la presencia de omisiones que reducen el sesgo en el análisis de los datos. Algunas de las metodologías de imputación han sido utilizadas comúnmente en la práctica a lo largo de los años por numerosos autores, pero otras están en el estado del arte con una menor utilización. Little y Rubin (2002) incluyen una taxonomía de métodos de imputación dependientes del tipo de problema y, siguen apareciendo nuevos procedimientos debido a que todo método de imputación sólo proporciona una respuesta parcial y, a veces, sesgada de los datos omitidos. Para la realización de proyectos de evaluación, como los casos de PISA, NAEP y Excale, se han utilizado modelos de imputación de datos faltantes con diversos grados de complejidad.

Los métodos para enfrentar las omisiones parten de un reconocimiento de su impacto, de modo que se acepta eliminar registros o hacer imputaciones hasta un cierto porcentaje de registros faltantes. En el caso de este estudio, con la base de datos de PISA 2006, se preparó el inventario de omisiones en las variables con el propósito de dictaminar la pertinencia de alguno de los métodos de imputación que se explican en la siguiente sección.

A2.2.3 Comentarios sobre los métodos de tratar las omisiones

En la metodología de trabajo para los modelos multinivel se puede proceder con dos grandes procedimientos empleados comúnmente:

- A) Eliminar de la base de datos los registros que tienen faltantes de la variable en estudio. Esto puede hacerse de varias formas: (a.1) directamente al momento de que el *software* de análisis multinivel prepara el archivo de estadísticas básicas, (a.2) por medio de un pre-proceso del archivo de datos, eliminando los registros con datos faltantes (íntegramente o con registros *ad hoc* para el análisis deseado¹²), (a.3) haciendo un pre-proceso del archivo de datos, produciendo una muestra controlada de los registros disponibles sin faltantes (en este caso se debe tener una muestra amplia, de tal modo que la muestra controlada tenga el número de elementos de diseño de acuerdo con el error planeado).
- B) Hacer imputación de los datos siguiendo modelos de asignación directa (por ejemplo la media o un valor que parte de una regresión), modelos de máxima verosimilitud (ML = *maximal likelihood*) o modelos de imputación múltiple (MI = *multiple imputation*), entre otras posibles formas sugeridas en la literatura.

La posibilidad de que el *software* elimine los registros que tienen omisiones no se consideró apropiada para el estudio; las recomendaciones internacionales explican que se pueden llegar a eliminar demasiados registros, con lo cual se haría poco útil el análisis; por otra parte se siguió la línea de otros trabajos del INEE (como es el caso de los análisis de la prueba Excale) donde se han hecho imputaciones en lugar de eliminar

¹² La eliminación de todos los registros faltantes se denomina *listwise* y el método de eliminar solo los registros que inciden en uno de los modelos jerárquicos se denomina *pairwise*.

registros. Esto brinda un mayor nivel de validez facial al estudio y, por ello, se adoptó para las consideraciones que se presentan a continuación.

Varios autores argumentan que la eliminación de registros no es aceptable porque puede producir sesgo en los análisis y una pérdida de poder estadístico (en el modo *listwise*). La eliminación de registros, dependiendo de la variable (*pairwise deletion*), es una práctica común en HLM, pero ha sido discutida como no recomendable en los análisis multivariados que requieren el uso de matrices de covarianza, lo cual es el caso de las ecuaciones estructurales, porque las poblaciones en cada análisis son diferentes al eliminar registros diferentes. La recomendación de APA, detallada por Thompson (1996), pide privilegiar a la imputación sobre eliminar registros en los estudios que se realicen desde 1994 a la fecha.

La imputación con valores centrados (como puede ser la media o la mediana) no se recomienda en general porque reduce la varianza real de los datos. Los modelos de regresión simple para asignar imputaciones sólo son operativos si se asigna alguna fluctuación estocástica a los datos, pero este tipo de fluctuación no es factible cuando se realizan modelos multinivel, debido principalmente a los efectos de anidamiento, por ello se requiere hacer imputaciones particulares por grupo lo cual, a su vez, afecta las estimaciones en los modelos jerárquicos. Los modelos ML y MI dependen del tipo de dato y del esquema de análisis, siendo MI el más general, en tanto que ML depende claramente del modelo a utilizar. Por ello, las imputaciones con ambos métodos son diferentes y deben implementarse con ayuda de *software* que realiza los procesos iterativos.

Una variante que se tiene disponible en el *software* de HLM consiste en hacer el análisis con múltiples imputaciones, para lo cual se preparan diferentes archivos de entrada al programa. En este caso, cada imputación se hace con dos posibles vertientes: (a) utilizar un modelo de imputación procedentes de un generador de números en condición totalmente aleatoria o por elección de una *semilla* diferente¹³, pero controlada por el analista; (b) utilizar diferentes modelos de imputación, en este caso se puede utilizar la misma *semilla* en todos los modelos o partir de un valor totalmente aleatorio. El uso de múltiples imputaciones no se consideró pertinente para los fines de este estudio, pero podría dar lugar a estudios complementarios en el futuro. Se eligió, en este caso, trabajar con un solo modelo de imputación¹⁴, como realizan Willms y Smith (2005).

¹³ La *semilla* es el valor de inicio para un algoritmo generador de números pseudo-aleatorios.

¹⁴ No se consideró de importancia hacer réplicas usando las ponderaciones de Fay contenidas en el archivo de PISA. Estas ponderaciones sólo se recomiendan si se desea calcular el error estándar del estadístico en estudio. En este caso, la variable dependiente de interés es el puntaje en Ciencias de la prueba PISA, el cual se traduce en los cinco valores plausibles, suficientes para el análisis y se pueden emplear directamente en HLM. Los estudios de la literatura no indican en forma explícita que han realizado modelos en tres niveles utilizando los valores plausibles de PISA, ni tampoco que utilicen las réplicas con las ponderaciones de Fay.

A2.2.4 La imputación en la base de datos de PISA

En el anexo 8 del reporte técnico de PISA 2006 (OCDE-PISA, 2007) se hace la observación de que, a nivel internacional, cada una de las variables tuvo una proporción menor a 5% de faltantes pero anotan que, por la distribución de las omisiones, si se realizara la eliminación de todos los casos se llegaría a una pérdida de muestra de 28.21%. Por ello decidieron realizar la imputación para conservar la mayor cantidad posible de casos en el análisis HLM.

Para el presente estudio, se imputaron datos a nivel individual y de escuela; cada imputación se acompaña de una variable *bandera* (variable muda o *dummy*), siguiendo el modelo propuesto por Cohen y Cohen (1983)¹⁵.

- A) Para las variables continuas se hizo la imputación con el promedio ponderado de la variable en la escuela; si este dato faltó se hizo la imputación con el promedio ponderado del país y, en caso de que este dato haya faltado, se utilizó el promedio ponderado internacional. La ponderación o peso empleado a nivel de escuela o de país es proporcional a la probabilidad de la muestra (factor de ponderación *W_FSTUWT* de la base de datos de PISA)¹⁶.
- B) Para las variables categóricas la identificación se hizo en función de una variable muda. Se asignó 1 si la observación pertenecía a la categoría y 0 en caso contrario.
- C) Para las otras variables mudas y dicótomas de la base de datos, los faltantes se imputaron con 0.

La ponderación en el nivel estudiante se hizo con los pesos normalizados finales (*W_FSTUWT*) de la base de datos PISA 2006, en dos niveles:

Nivel país para análisis de regresión de dos niveles. La suma de los pesos dentro de cada país se hace que sea igual al número de estudiantes en la base de datos, o sea, igual al tamaño muestral de esa nación y se mantiene la misma proporción que tenga el peso final de los alumnos dentro del país (*W_FSTUWT*).

Nivel internacional para análisis de regresión de tres niveles. Los pesos finales se normalizaron con 55 de los 57 países participantes, para que la suma de pesos fuera igual al número de estudiantes en los 55 países de la base y se mantuviera la misma proporción de los pesos finales de los estudiantes dentro de cada país (*W_FSTUWT*). El procedimiento garantiza que las contribuciones de los países individuales sean iguales utilizando un factor de país, de este modo la suma de los pesos dentro de cada país es igual para todos los países del estudio.

¹⁵ El modelo de Cohen y Cohen puede presentar un problema de sesgo que se ha considerado como menor.

¹⁶ Los ponderadores internacionales usaron iguales pesos por país referido a mil casos.

A2.2.5 La imputación relacionada con los programas de análisis multinivel y de ecuaciones estructurales

Raudenbusch, S., Bryk, A. y Congdon, R. (2005) indican¹⁷ que el manejo de los datos faltantes debe ser apropiado a cada caso y es responsabilidad del usuario. Para los modelos de 2 y 3 niveles (HLM2, HLM3), el *software* no hace imputación sino que elimina directamente los datos faltantes al momento de generar el archivo de estadísticas básicas (archivo MDM) o durante la corrida. La eliminación en la etapa de generación del archivo MDM se basa en las variables elegidas del nivel 1, si la eliminación se hace en la corrida de análisis, la eliminación se efectúa solamente en las variables que intervienen en el modelo. Los otros modelos multivariados de HLM no permiten el uso de datos faltantes.

Cuando se desea realizar la imputación, el modelo de asignación de los datos faltantes es responsabilidad del usuario al momento de preparar los datos, observando que si se dejan en blanco los campos de los datos faltantes en el archivo, serían tratados como respuestas válidas en valor nulo.

Arbuckle, J. y Wothke, W. (1999) indican¹⁸ que las soluciones típicas para imputación corresponden con la eliminación de casos (*listwise* y *pairwise*) o la asignación de la media. Se dice que cuando hay menos de 5% de casos faltantes (al azar), la eliminación de registros es aceptable, pero sugieren revisar que las omisiones sean aleatorias para justificar la eliminación ante posible sesgos inducidos en la eliminación de registros.

El *software* AMOS no hace imputaciones, pero utiliza los datos disponibles con un modelo de máxima verosimilitud (siguiendo a Little y Rubin, 2002), que es el modelo de imputación más recomendado, especialmente cuando los datos faltantes son aleatorios y producen el menor sesgo en estudios multivariados.

A2.2.6 Imputación para los datos de omisión en este estudio

A diferencia de otros estudios, como es el caso de los análisis realizados para Excale donde se hace una asignación por un modelo de regresión o por medio de la mediana, para este estudio se decidió emplear el método *HotDeck*, o de asignación modal por donantes similares a los casos faltantes, junto con la asignación de una variable muda, similar al modelo propuesto por Cohen y Cohen (1983) y presentado posteriormente por Allison (2002), McKnight y col. (2006).

El modelo de donantes asigna la media del conjunto de datos más similares al registro faltante, de acuerdo con el procedimiento empleado en el reporte internacional de PISA. Se consideran subgrupos de donantes similares, por ejemplo por tipo de escuela del mismo nivel socioeconómico u otra característica dentro de un estado, con lo cual se construyen estratos para las variables continuas y donantes de un solo valor para las variables discretas¹⁹. Una vez hecha la asignación, se construye

¹⁷ Sección *Handling of missing data in HLM* de la guía de ayuda del programa HLM

¹⁸ Sección *Complete Data or Appropriate Handling of Incomplete Data*

¹⁹ Se propuso también considerar la generación de por lo menos tres valores aleatorios para cada subgrupo y después calcular su media, pero no se obtuvo ninguna ventaja en contraste con el tiempo de desarrollo de los cálculos para todo el conjunto de datos faltantes.

una variable muda (variable *dummy* o *bandera*) dicótoma, donde se indica con el número 1 si el caso fue imputado.

Este método combina las ventajas del modelo de imputación de donantes o *HotDeck* con el escalamiento que produce la variable muda y evita el uso de modelos de simulación, los cuales pueden servir para otros fines pero no para este caso, donde los datos y su distribución son desconocidos de antemano; de hecho un proceso de simulación tiene sentido cuando los valores se pueden contrastar con un modelo de referencia o con los datos de los cuales forma parte la muestra; para el caso de los datos de PISA 2006, la simulación sólo permitiría contrastar con los mismos datos, lo cual carece de utilidad. A continuación se describen los procedimientos para imputar los valores omitidos en las bases de datos de PISA 2006.

El procedimiento para imputar valores perdidos en los índices, a nivel estudiante, es el siguiente:

1. Seleccionar el índice que presenta valores perdidos.
2. Crear una variable *bandera* (variable muda o *dummy*) de tipo binario para identificar los casos en que ocurren las omisiones. A la variable se le asigna un valor de 1 si debe imputarse por estar omitida y un valor de 0 en otro caso.
3. Calcular el promedio del índice a imputar para cada escuela.
4. Asignar el promedio de la escuela a cada estudiante que presenta valores perdidos.
5. Repetir el mismo procedimiento para cada índice con valores perdidos.

Este es el procedimiento para imputar valores perdidos en los índices a nivel escuela:

1. Seleccionar el índice que presenta valores perdidos.
2. Crear una variable *bandera* (variable muda o *dummy*) de tipo binario para identificar los casos donde se imputan los valores omitidos, asignando 1 si el valor está perdido y 0, en caso contrario.
3. Crear grupos de estudiantes de acuerdo con las siguientes variables de clasificación: el estado donde se encuentra (entidad federativa), el nivel educativo al que asisten (Secundaria y Educación media superior), el tipo de sostenimiento de la escuela (Pública o Privada) y la comunidad donde se encuentra la escuela (Urbana o Rural).
4. Promediar el índice de los estudiantes en cada grupo.
5. Asignar a cada escuela el promedio del estrato al que pertenece con valores perdidos.
6. Si el valor perdido del índice a imputar se encuentra en un estrato donde no hay información, entonces:
 - a) Calcular el promedio del índice para todos los estudiantes de la entidad federativa.
 - b) Asignar a la escuela el promedio calculado por entidad federativa.
7. Repetir el mismo procedimiento para cada índice de las escuelas donde hay omisiones y se deben imputar valores.

A2.3 Sobre los pesos o ponderaciones individuales y regionales

En muestras como la que se emplea en el proyecto PISA se acostumbra asignar pesos muestrales que dimensionan la aportación de cada uno de los sujetos con lo que se corrigen las deficiencias que llegaran a presentarse al establecer el tamaño muestral o durante la recopilación de datos. Dentro de los métodos para corregir la asignación de peso se destaca el procedimiento de corte o *trimming*, que puede efectuarse por escuela, por estado o por región.

El método de *trimming* parte de considerar las ponderaciones de diseño para las escuelas de un estado por estrato (por ejemplo en función del número de estudiantes se clasifican las escuelas en grandes, medias y pequeñas). Puede ocurrir que la suma de las ponderaciones asignadas exceda el tamaño representativo de la población del estado, lo cual obliga a ajustar los ponderadores por medio de un recorte (*trim*), quedando un remanente que debe redistribuirse en todo el estrato estatal. Como resultado del proceso se espera llegar a unos valores marginales de género, Público-Privado, Urbano-Rural, Media superior-Secundaria.

Las ponderaciones o pesos que aparecen en la base de datos de PISA están estandarizados y se utilizan en el *software* directamente al momento de calcular los modelos multinivel. Por lo anterior, no se hizo ninguna recalibración de los pesos muestrales por región.

El problema de ponderar o asignar pesos a las variables, especialmente en el caso de los modelos multinivel, estriba en la variedad de posibilidades para hacer la asignación de pesos. Por ejemplo en el documento de ponderación y muestreo del proyecto PISA (OECD-PISA, 2003) se menciona que las ponderaciones deben ser asignadas por el propio Proyecto con réplicas balanceadas o con el modelo de muestreo *jackknife*²⁰, pero no se hace mención a que puedan presentarse problemas de asignación de pesos en el análisis multinivel. Ponderaciones similares a las de PISA (no para análisis multinivel sino solo para estudiar el funcionamiento diferencia de los ítems y de las pruebas) se han usado en otros proyectos, como el caso del IAEP (Chu y col., 1992), donde se diseñaron muestras de 30 a 35 estudiantes por escuela; cada muestra se pondera para hacer referencia al país, llegándose aproximadamente a 1650 estudiantes por muestra de cada país (en dicho proyecto participaron China, Unión Soviética, Canadá y sus provincias, España, Jordania, entre otros más). Similares ideas son proporcionadas por Mullis y col. (1993) para la prueba NAEP.

Durante el desarrollo de los análisis se tuvieron algunos problemas de convergencia que obligaron a rediseñar las ponderaciones de PISA. El primer paso consistió en redimensionar los ponderadores originales que representaban a la población objetivo y el segundo, diseñar otros ponderadores estandarizados (o índice por regiones), que son los que se emplean en los cálculos.

²⁰ El método *jackknife* es un procedimiento para construir sub-muestras a partir de una muestra dada, con el objeto de tener que calcular el estándar de los estimadores estadísticos, aunque inicialmente fue utilizado para calcular sesgo.

A2.3.1 Cálculo de las ponderaciones de los estudiantes

El diseño muestral se estableció en 35 estudiantes para tener una representatividad a nivel de escuela. Este número de personas no es proporcional al número de alumnos de la escuela, de la región o del estado, por lo tanto se vuelve necesario considerar un ponderador para cada estudiante a manera de reportar resultados representativos a nivel regional.

Los pesos o ponderadores que se asignan a cada estudiante para que representen al número de alumnos en la población pueden ser a nivel escolar (proporcionados directamente en la base de datos de PISA con la etiqueta W_FSTUWT) o a nivel región como se explican a continuación. Para el análisis regional multinivel se consideraron los datos ponderados y normalizados a nivel estudiante. El objetivo de la normalización es dar a cada región el mismo peso para evitar el sesgo por tamaño de la población. La normalización de los pesos se realizó de la siguiente manera:

1. Se tomó la base de datos de estudiantes de PISA 2006 y se suprimieron aquellos inscritos en escuelas secundarias para trabajadores y capacitación para el trabajo.
2. Se asignó a cada alumno la región que le correspondía de acuerdo con la ubicación de la escuela en la que estaba inscrito.
3. Se realizó la suma de los pesos finales (W_FSTUWT) de los estudiantes para las siete regiones en las que se dividió el país.
4. Se igualó la suma de los pesos al número de estudiantes en la muestra en cada región.
5. Se multiplicó el número de alumnos en la región por el cociente del ponderador (W_FSTUWT) entre la suma de ponderadores en la región.
6. Se dividió el número total de estudiantes en la base de datos entre siete y el resultado obtenido se dividió entre el número de escolares de cada región.
7. Finalmente, se multiplicó el cociente del paso (4) con el del paso (6) para obtener el peso normalizado para regiones, que es el que se emplea en el *software* HLM para la ponderación de cada estudiante.

Siguiendo este procedimiento se tiene una suma de pesos constante para las siete regiones.

A2.3.2 Cálculo de los índices de regiones

Los índices socioeconómicos, demográficos y educativos del nivel 3 se encuentran disponibles en las referencias para las entidades federativas, por lo que se tuvo la necesidad de recalcularlos para cada región. El promedio simple de los índices de las entidades agrupadas en cada región generaría sesgo, pues las entidades con mayor población estarían subrepresentadas con el índice calculado (denominado St_WG-TRE). Para corregir esta desviación los índices de las regiones se calcularon ponderando los índices de cada entidad de la siguiente manera:

1. Se obtuvo el tamaño de la población en cada entidad, tomando como referencia el año 2005.
2. Se sumó la población de todas las entidades agrupadas en cada región, obteniendo un total para cada región.
3. Se dividió la población de cada entidad entre el total de la región a la que pertenece, obteniéndose, de esta manera, el factor de ponderación.
4. Se calculó el producto del factor de ponderación por el índice o por la variable de interés en cada entidad.
5. El índice recalculado de cada región es la suma de los productos del factor de ponderación por el índice para las entidades en la región.

A2.4 Sobre los modelos multivariados de ecuaciones estructurales

Una de las recomendaciones más extendidas en relación con estudios similares al realizado en este trabajo, enfatiza la necesidad de utilizar modelos multivariados para modelar los datos. Los textos de referencia como el de Morrison (1967) proporcionan los fundamentos de los métodos de análisis factorial, así como las bases del desarrollo de modelos de regresión multivariados; de estos modelos a la utilización de ecuaciones estructurales no hay más que un paso. Se puede partir de modelos de trayectoria o de senderos (*path analysis*), que se enfocan a analizar variables observadas (por ejemplo, Nunnally y Bernstein, 1995, Raykov y Marcoulides, 2000), pero actualmente se cuenta con *software* muy eficiente que permite incluir variables latentes, con las cuales se desarrollan las denominadas ecuaciones estructurales. Este enfoque se puede asimilar a modelos factoriales confirmatorios, que se plantean como alternativa o complemento de los estudios multinivel, con el propósito de confirmar hipótesis y proporcionar un soporte adicional al estudio de las variables del problema. Las ecuaciones estructurales proporcionan una base para explicar relaciones causales entre las variables.

Se hicieron modelos con ecuaciones estructurales para identificar variables explicativas de los constructos latentes del Modelo de Calidad Educativa del INEE. Este análisis debe considerarse sólo como preliminar: no se pretende confirmar ninguna hipótesis de causalidad entre las variables propuestas porque debe recordarse que, en este momento del desarrollo, ni los cuestionarios de contexto de la prueba PISA, ni los datos socioeconómicos utilizados en el tercer nivel, están considerando los ejes del Modelo de Calidad. Los modelos estructurales deben ser vistos como un apoyo a la comprensión de la justificación del marco teórico y como un complemento del análisis factorial exploratorio realizado para este estudio. Los aspectos rescatables de estos modelos no son los valores crudos que se obtienen de las correlaciones entre variables observadas, ni tampoco las relaciones causales con las variables latentes, sino la proporción relativa que tienen los coeficientes asociados a cada eje y, sobre todo, su signo.

Estos análisis multivariados proporcionan una visión cualitativa-cuantitativa de la forma en que conjuntos de variables pueden estar asociados con los resultados de la prueba PISA, en forma directa o a través de variables latentes mediadoras. Con los

resultados obtenidos se tiene una primera aproximación de las dimensiones asociadas con el Modelo de Calidad Educativa del INEE.

Los modelos que se construyeron con las variables de la base de datos de PISA partieron de la información proporcionada por un análisis factorial exploratorio. Cada análisis pudo relacionarse con alguna de las dimensiones del modelo de calidad y con ello se confirmó la presencia del factor por un lado, y la posible interacción de las variables dependientes e independientes por otro lado. No se trata de modelos que involucran a los diferentes niveles de anidamiento (estudiantes, escuelas, estados, regiones), lo cual se hará con los modelos multinivel; en cambio se construyeron como modelos factoriales confirmatorios que resultan interesantes porque indican la proporción de participación de cada variable y el signo, que indica el sentido de la variable explicativa dentro del comportamiento de las otras variables.

Para el presente estudio se eligió el *software* AMOS 5 (Arbuckle, 2003), que permite calcular los coeficientes estandarizados del sistema de ecuaciones de regresión que asocian variables observadas y latentes, tomando en cuenta también los componentes de errores de las varianzas. El modelo estructural, además de identificar las relaciones entre variables se emplea como análisis factorial (confirmatorio y descriptivo) de la asociación que guardan entre sí las variables de cada nivel de anidamiento.

A continuación se presenta un ejemplo y la forma de interpretar los resultados del análisis. Se incluye la representación gráfica de un modelo, sin entrar en detalles de las ecuaciones establecidas entre las variables ni los aspectos de ajuste y convergencia.

Las ecuaciones estructurales pueden partir de la matriz de varianzas-covarianzas de las variables de interés, las cuales se relacionan en forma directa como variables observadas o a través de variables adicionales que se definen como latentes. Estas relaciones se escriben como ecuaciones cuyas incógnitas a calcular son los coeficientes que multiplican a las covarianzas de las variables seleccionadas para el modelo. El conjunto de ecuaciones en términos de los coeficientes no son lineales, para su solución requieren de técnicas numéricas que obligan al uso de la computadora. La construcción misma de las ecuaciones es complicada, por ello el *software* ofrece un medio ambiente gráfico para poder construirlas de forma práctica.

De manera convencional, en el modelo gráfico las variables latentes se representan por medio de elipses, y las variables observadas por rectángulos. Sobre cada variable se puede hacer interactuar un componente de error (representado por un círculo pequeño que apunta hacia la variable)

Las flechas que ligan a las variables latentes u observadas permiten definir el sentido de construcción de las ecuaciones estructurales, con lo cual se calculan los coeficientes que establecen la forma en que las variables dependientes (denominadas variables explicativas) son influidas por variables independientes (observadas o latentes). Una vez que se diseña el modelo, el programa

resuelve el sistema no lineal de ecuaciones, proporcionando la relación entre las varianzas y covarianzas de las variables, la fracción de error de medida en la variable y los coeficientes estructurales (similares a coeficientes de regresión) que pueden presentarse en forma normalizada o no (en términos de covarianzas o de correlaciones). Los valores que aparecen a un costado de las flechas proporcionan coeficientes que ilustran la medida de la contribución de la variable.

Las relaciones entre variables no son necesariamente causales, pero por la forma en que se establecen las ecuaciones se dice que son formas que *explican* las varianzas entre variables independientes y dependientes.

Un modelo estructural no representa la única solución posible para el conjunto de relaciones que pueden establecerse entre las variables. Se trata de opciones *explicativas* que dependen del criterio del diseñador y de la información previa disponible, aunque puede haber otros modelos que también produzcan un buen conjunto de ecuaciones e asociación entre las variables produciendo expresiones *explicativas* de diferente valor. Debido a la limitante de que el modelo estructural no es único y a que los datos están anidados por escuela y por región (lo cual no se incluye en estos análisis), se debe insistir que sólo se emplea cada modelo para confirmar los análisis factoriales y como análisis preliminar del modelo multinivel desarrollado en el proyecto.

Los modelos estructurales pueden ser divergentes si las variables están altamente correlacionadas entre sí, o escasamente relacionadas con la variable dependiente. También pueden fallar si se tiene una pequeña cantidad de unidades de información (sujetos, escuelas o registros). El *software* proporciona varios posibles índices de ajuste, debido a la oferta de esquemas que hay en la literatura especializada y a la divergencia de opiniones de los investigadores a este respecto. Para los análisis realizados aquí se decidió utilizar principalmente el índice de bondad de ajuste AGFI que es una medida de la cantidad de varianza y covarianza del modelo real explicado por el modelo teórico y normalizado respecto del número de grados de libertad calculados (Byrne, 2001). Un buen ajuste se obtiene con valores cercanos a 1, en este estudio se consideran aceptables los valores superiores a 0.8. En algunos casos de la base regional, el modelo no es convergente por estar basado en siete registros únicamente, entonces se decidió utilizar el índice comparativo de ajuste CFI, que también debe ser muy cercano a 1 para poder afirmar que hubo ajuste del modelo.

Para los propósitos de este estudio no se tomaron los valores de ajuste en forma estricta; debe recordarse la limitación en cuanto a registros disponibles y a que las ecuaciones estructurales se realizaron como análisis factorial confirmatorio en un solo nivel, sin mediar la inclusión de los niveles de anidamiento.

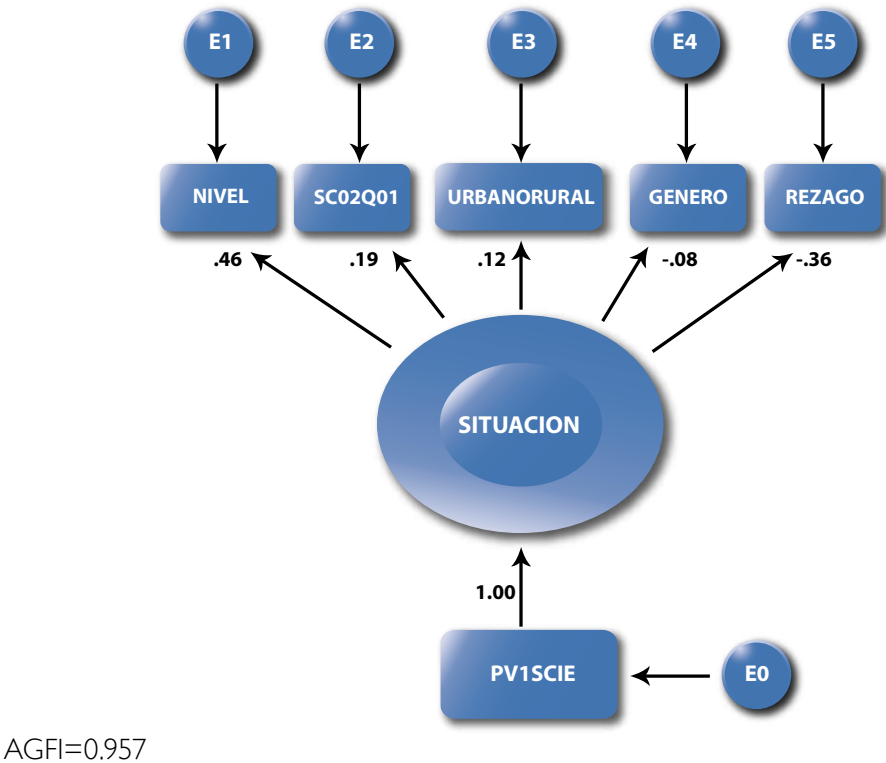
Ejemplo e interpretación

Se ilustra el uso de las ecuaciones estructurales con el modelo de la situación del estudiante. Las variables elegidas de la base de datos del primer nivel son:

Tabla A2.5 Variables del ejemplo de ecuaciones estructurales

Nivel	Variables	Descripción
Variables de nivel I	NIVEL	Nivel de estudios del estudiante: 0=Secundaria, 1=Educación media superior
	SC02Q01	Tipo de sostenimiento de la escuela: 0=Público, 1=Privado
	URBANORURAL	Tipo de comunidad donde se encuentra la escuela: 0=Rural, 1=urbano
	GENERO	Género del estudiante. Codificación: 0=hombre, 1=mujer.
	REZAGO	Se considera que un estudiante es rezagado si cumplió años en el periodo de enero-agosto de 1990 y está inscrito en secundaria. Codificación: 0=Sin rezago, 1=Rezagado

Figura A2.3 Modelo de ejemplo de ecuaciones estructurales



El modelo de la figura A2.4 indica que la variable latente denominada SITUACION se divide u organiza en cinco variables explicativas observadas de la base de datos de PISA. A su vez la variable observada del desempeño en Ciencias debe verse reflejada en la SITUACION. De las variables asociadas con SITUACION se observa que NIVEL y tipo de escuela SC02Q01 (Pública o Privada), son las que se asocian de manera más notable en sentido positivo (sus coeficientes son 0.46 y 0.19), en tanto que REZAGO es la más importante en sentido negativo (-0.36). Este modelo de ecuaciones estructurales puede interpretarse afirmando que a un estudiante el rezago le produce un efecto adverso en su desempeño en Ciencias ($REZAGO=1$); también se puede afirmar que un estudiante de escuela privada ($SC02Q01=1$) tiene mayor desempeño que el de una escuela pública ($SC02Q01=0$). El género del alumno ($Mujer=1$) también tiene un efecto negativo en el desempeño en Ciencias. Todas estas variables inciden en el desempeño en Ciencias con un coeficiente de 1 (similar a una regresión entre las variables). El ajuste es bueno con un valor AGFI=0.957.

A2.5 Modelos multinivel

Los modelos jerárquicos lineales se han vuelto una exigencia para poder analizar datos procedentes de exploraciones en que las unidades de estudio no son independientes, sino que están anidadas o agrupadas, implicando una correlación entre los elementos de cada grupo (Hox, 1995; Kreft y DeLeeuw, 1998; Goldstein, 1999). Es evidente que la muestra de estudiantes no es estrictamente aleatoria: los alumnos de una escuela correlacionan mucho entre sí porque se ven influidos por factores grupales, de aula y de la escuela; las escuelas de un estado correlacionan mucho más entre sí que con las escuelas de otro estado.

Debido al anidamiento, los rasgos asociados con los sujetos presentan una alta correlación intragrupal que no favorece el uso de técnicas de regresión tradicionales, porque se producen correlaciones más altas de lo esperado, siendo preferibles los modelos multinivel que permiten desglosar o separar la dependencia por nivel de los datos en partes fijas (interceptos y coeficientes de regresión) y aleatorias (errores o diferencias entre los datos y el resultado teórico que se obtiene con los valores fijos) (Gaviria, Martínez y Castro, 2004). De esta manera se pueden reportar las participaciones de cada variable y cada junto con la varianza intra-grupal e inter-grupal simultáneamente.

Al diseñar un modelo multinivel se identifican por una parte, los tipos de anidamiento en los cuales se agrupan los datos y, por otra parte, las variables relacionadas con la variable dependiente (denominadas variables explicativas, pero que no representan relaciones de causa-efecto con la variable dependiente). Dada la complejidad del cálculo, se requiere de *software* especializado y de las capacidades actuales de las computadoras, habiendo actualmente una gran oferta de programas para esta especialidad. Para este estudio se eligió el *software* HLM6.0 (Bryk, S., Raundebush, A. y Congdon, R., 2005).

Hay un buen número de ventajas con los modelos multinivel, entre ellas que el modelo reporta las relaciones entre las variables de los diferentes niveles, corrige el

efecto de autocorrelación que se tiene por el anidamiento de los datos y favorece la identificación del impacto de cada nivel en el modelo de predicción entre las variables explicativas y la variable dependiente.

Por todo lo descrito, una investigación sobre la calidad educativa con base en el marco teórico que fundamenta este estudio, se realiza convenientemente con modelos multinivel, porque permiten partir de los aspectos recogidos a través de los cuestionarios contextuales de PISA para responder a las inquietudes relativas a la definición de algunos de los aspectos que forman a los ejes de calidad e identificar variables que se relacionan con incrementos o decrementos en el desempeño en Ciencias de los estudiantes.

En relación con los niveles de anidamiento, la revisión de la literatura muestra que su elección no es motivo de grandes discrepancias. Queda claro que en este estudio se consideran tres niveles de anidamiento: Estudiante-Escuela-Región y la variable dependiente es el puntaje global de Ciencias, sin considerar sus subescalas.

En cambio, las variables explicativas que se emplean en los modelos no están definidas de antemano, aunque existan diversas corrientes o propuestas para su elección. Para este estudio se prefirió partir del marco teórico sobre la calidad educativa en lugar de repetir la elección de variables de alguno de los estudios multinivel sobre eficacia educativa que hay en la literatura. El enfoque seguido para este estudio acota la participación de los diferentes índices o variables que permiten explicar cada una de las dimensiones del modelo de calidad. A partir del marco teórico, las variables explicativas que se emplean en este estudio pueden diferir parcialmente de las que se han utilizado en otros estudios, especialmente porque en algunos estudios dicho marco de referencia es sustancialmente diferente o, sobre todo, porque en otros estudios no existe marco alguno. Así, por ejemplo, en un estudio realizado en España (Ruiz y Castro, 2006), se consideran estas variables en los diferentes niveles de anidamiento para un estudio sobre la eficacia educativa (obsérvese que es una sola de las dimensiones del Modelo de Calidad Educativa del INEE), las cuales tienen ciertas semejanzas y diferencias con las utilizadas en el presente estudio.

Nivel 1: estructura familiar, sexo, nivel socioeconómico, recursos educativos en el hogar, expectativas educativas del estudiante.

Nivel 2: nivel socioeconómico medio de la escuela, tamaño de la escuela, titularidad de la escuela (Público-Privado), calidad de los recursos educativos, proporción de profesores con certificación pedagógica.

Nivel 3: todos los países, porcentaje del PIB, la inversión del país en la educación secundaria.

Una vez elegidos los niveles y las variables pertinentes, se construye el modelo multinivel que es un conjunto de ecuaciones lineales que expresan la relación entre variables independientes y la variable dependiente. Los coeficientes de la ecuación en uno de los niveles se asocian, a su vez, con las variables del nivel siguiente por medio de otro conjunto de ecuaciones lineales. Los coeficientes se denominan interceptos (si son independientes) o pendientes (si multiplican a una variable). Las diferencias entre los valores reales observados y los teóricos que se obtienen con el modelo, son aleatorias pero se hace la hipótesis de que tienen una distribución normal de media

nula y una cierta varianza que el programa reporta en cada nivel por separado. Se debe verificar que el modelo converge, a través de la revisión de la varianza total del modelo; el modelo es divergente o inválido si el error estándar de la varianza total no puede calcularse o si es negativo. Cuando se identifica un modelo divergente se tienen dos opciones: (1) proponer una nueva combinación de variables explicativas en el nivel de interés o (2) eliminar del modelo las variables que tengan una participación prácticamente nula en la explicación de la varianza total.

Como ejemplo de análisis multinivel se presenta el modelo de la situación del estudiante, donde se eligieron variables del primer nivel (tabla A2.6).

Tabla A2.6 Variables del ejemplo de modelos multinivel

Nivel	Variables	Descripción
Variables de nivel I	NIVEL	Nivel de estudios del estudiante: 0=Secundaria, 1=Educación media superior
	SC02Q01	Tipo de sostenimiento de la escuela: 0=Público, 1=Privado
	URBANORURAL	Tipo de comunidad donde se encuentra la escuela: 0=Rural, 1=urbano
	GENERO	Género del estudiante. Codificación: 0=hombre, 1=mujer.
	REZAGO	Se considera que un estudiante es rezagado si cumplió años en el periodo de enero-agosto de 1990 y está inscrito en secundaria. Codificación: 0=Sin rezago, 1=Rezagado

La ecuación del modelo multinivel es:

$$PVSCIE_{ijk} = \gamma_{000} + \gamma_{100} NIVEL_{jk} + \gamma_{200} SC02Q01_{ijk} + \gamma_{300} URBANORURAL_{ijk} + \gamma_{400} GENERO_{ijk} + \gamma_{500} REZAGO_{ijk} + r_{0jk} + u_{00k} + e_{ijk}$$

En esta expresión los tres subíndices ijk de cada coeficiente o variable se refieren a los tres niveles de anidamiento. Se distinguen los denominados términos fijos, que son comunes para todos los sujetos independientemente del nivel en el que se encuentran: en primer lugar se tiene γ_{000} , denominado *intercepto*, que representa un valor base para el cálculo de la puntuación en Ciencias de PISA para un estudiante, y los coeficientes de las variables (como γ_{100} , γ_{200}) que detallan la forma en que un cambio unitario en la variable explicativa modela el desempeño en Ciencias.

El valor e_{ijk} representa la fluctuación que tiene la puntuación real del estudiante respecto del modelo teórico que se establece con el intercepto y los términos de las variables, en el nivel I de anidamiento. La intervención del anidamiento ESCUELA queda incluida en el segundo nivel a través del coeficiente aleatorio r_{0jk} . No se incluye explícitamente el anidamiento REGION, que queda contenido en el tercer nivel y cuya contribución se aprecia en el coeficiente aleatorio u_{00k} .

Al término de la corrida con el programa HLM6 se obtienen el intercepto y los coeficientes, o términos fijos y las varianzas de los coeficientes aleatorios (tabla A2.7).

Tabla A2.7 Términos fijos del modelo multinivel de las variables de situación del estudiante

Efecto fijo		Coefficiente	Error estándar	t	Grados de libertad	p-valor
Interceptos						
Puntuación base global en Ciencias	γ_{000}	371.57	8.54	43.51	6	0.000
Pendientes						
NIVEL	γ_{100}	44.97	3.71	12.12	243	0.000
SC02Q01	γ_{200}	29.32	6.09	4.82	2018	0.000
URBANORURAL	γ_{300}	32.96	5.28	6.25	65	0.000
GENERO	γ_{400}	-15.86	1.72	-9.23	4375	0.000
REZAGO	γ_{500}	-14.95	2.94	-5.09	54	0.000

Si se substituyen estos coeficientes o términos fijos en la ecuación del modelo multinivel, se llega a la expresión:

$$PVSCIE_{ijk} = 371.57 + 44.97 NIVEL_{jk} + 29.32 SC02Q01_{ijk} + 32.96 URBANORURAL_{ijk} - 15.86 GENERO_{ijk} - 14.95 REZAGO_{ijk} + r_{0jk} + u_{00k} + e_{ijk}$$

Con ayuda de los coeficientes se puede determinar la influencia de cada variable. Cuando las variables son dicótomas (como REZAGO, GENERO, URBANORURAL) el cambio en la variable se calcula multiplicando el coeficiente por el paso de una categoría a otra; cuando se trata de variables continuas (como el estatus socioeconómico ESCS) el cambio debe verse como relacionado con un incremento de una unidad en la variable (como cuando se pasa de ESCS=0.5 a ESCS=1.5). Por ejemplo, un estudiante rezagado (REZAGO=1) reduce 14.95 (-14.95×1) su puntuación de PISA respecto del valor base dado por el intercepto; en cambio un estudiante que está en Educación media superior (REZAGO=0) no se ve afectado por este coeficiente (-14.95×0).

Todos los coeficientes cuentan con un error estándar que puede utilizarse para definir intervalos de aceptación, generalmente con 95% de confianza (para fines prácticos, el intervalo abarca dos errores estándar por arriba y por debajo del valor nominal obtenido). En este caso, el límite superior del intercepto estaría en 388.65 ($371.57 + 2 \times 8.54$) y el límite inferior en 354.49 ($371.57 - 2 \times 8.54$), con lo que se concluye que el intercepto no es nulo ni se encuentra cerca de cero.

Para determinar de forma sistemática que un coeficiente es significativo para el modelo, el programa realiza la prueba de hipótesis de nulidad con el valor de t ($t = \text{Coeficiente} / \text{Error estándar}$), que debe ser significativamente diferente de cero ($p \leq 0.05$, para el número de grados de libertad implicados en el cálculo). Todos los coeficientes del análisis mostrado tienen valores significativamente diferentes de cero, pero si llegara a tenerse un valor estadísticamente nulo el diseñador podría eliminarlo del modelo.

La tabla A2.8 engloba los resultados de la varianza de los términos aleatorios que corresponden con las fluctuaciones que se tienen para cada estudiante de acuerdo con el modelo teórico obtenido al incluir los coeficientes fijos. Estas fluctuaciones se distribuyen en forma normal con media 0; su desviación estándar se proporciona en la tabla.

Tabla A2.8 Términos aleatorios del modelo multinivel de la situación del estudiante

Efecto aleatorio			Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
Nivel	Unidad							
1	Estudiante	e_{ijk}	57.76	3336.42				61.00
2	Escuela	r_{0jk}	44.33	1965.28	1081	13841.11	0.0	35.93
3	Región	U_{00k}	12.95	167.64	6	89.71	0.0	3.07

La varianza total es la suma de los componentes de varianza para cada nivel, información que solo se obtiene con los modelos multinivel y que los modelos de regresión clásicos no pueden separar de manera automática. La mayor varianza ocurre en el nivel 1 (por influencia del desempeño de los estudiantes), ocupando 61% de la varianza total; el segundo nivel ocupa 35.93% de la varianza total, en tanto que el tercer nivel solo corresponde con 3.07% de la varianza total. El *software* proporciona también una prueba de hipótesis para verificar que estos componentes de varianza pueden considerarse estadísticamente diferentes de cero, por medio de una prueba χ^2 , para los grados de libertad involucrados en el análisis.

Hasta aquí se tienen los resultados procedentes del *software* HLM, pero es necesario analizar los coeficientes fijos y aleatorios para dictaminar si las categorías de las variables en estudio producen diferencias significativas en el desempeño de los estudiantes en la prueba de Ciencias. Este análisis se presenta en el anexo A2.8.

A2.6 Sobre el uso de variables bandera en los modelos multinivel

En las secciones A2.2.2 a A2.2.4 se explicó el modelo de imputación que se utilizó en este estudio, con base en las recomendaciones de otros estudios similares, asignando un valor de índice faltante por el método de donantes o *HotDeck* y la inclusión de la variable *bandera* que permite guardar una referencia de los casos en los que se hizo la imputación.

Cada vez que hay una variable imputada, en la base de datos se puede encontrar una pareja de campos asociados: el campo de la variable propiamente dicha y una variable muda asociada, donde se refiere si dato es original o es imputado. Por ejemplo, se tienen los campos ESCS y BaESCS, donde la pareja (-1.4, 0) indica que el nivel socioeconómico y cultural de la persona es -1.4 y el dato no es imputado, en tanto que la pareja (0.6, 1) indica que el dato fue imputado con el valor 0.6 en función del subgrupo de personas más afines.

La presencia de la variable muda permite no solamente identificar que se trata de una imputación, sino contar con un elemento que dimensiona o ajusta el coeficiente de regresión asociado con la variable. Este empleo de las variables mudas ha sido propuesto por Cohen y Cohen (1983) y Willms y Smith (2005)²¹, quienes apuntan que el método es de aplicación práctica, aunque debe tenerse cuidado por el uso excesivo de variables mudas que se pueden requerir en un modelo, así como la posibilidad de que se tengan estimaciones sesgadas. A pesar de estas posibles deficiencias, Willms utiliza la substitución con la media y el uso de la variable muda para los análisis que realiza con TIMMS y PISA.

El uso del modelo de imputación por donantes *HotDeck*, por subgrupos como se sugiere aquí, combinado con la variable muda no se ha explorado en las referencias, por lo que se consideró importante analizar la influencia de la variable muda para cada modelo. Para la validación del modelo se hicieron dos procedimientos: el primero consistió en obtener las rectas de regresión simple del primer valor plausible del puntaje de Ciencias, respecto del índice socioeconómico y cultural ESCS; esta regresión se efectuó con la variable muda y sin ella, para poder cotejar la diferencia entre los coeficientes. El segundo procedimiento extiende el paso anterior, utilizando una predicción con modelos multinivel, en dos casos particulares:

- a) Modelo que incluye una variable explicativa y su *bandera* para el primer nivel y un modelo incondicional (sin variables) en los niveles 2 y 3.
- b) Modelo que incluye una variable explicativa y su *bandera* para el segundo nivel, con modelo incondicional (sin variables) en los niveles 1 y 3.

Esta verificación permite observar el efecto que se tiene en un cambio cuantitativo del ajuste del modelo el cual, a su vez, se refleja en una reducción del desajuste (*deviance*) del modelo. Por ejemplo, se construye un primer modelo con la variable HOMEPOS y un segundo modelo incluyendo también su variable *bandera* asociada, como se indica a continuación:

Modelo I	Modelo II
Level-1 Model $Y = P0 + P1*(HOMEPOS) + E$	Level-1 Model $Y = P0 + P1*(HOMEPOS) + P2*(BAHOMEPO) + E$
Level-2 Model $P0 = B00 + R0$ $P1 = B10$	Level-2 Model $P0 = B00 + R0$ $P1 = B10$ $P2 = B20$
Level-3 Model $B00 = G000 + U00$ $B10 = G100$	Level-3 Model $B00 = G000 + U00$ $B10 = G100$ $B20 = G200$

²¹ El método original propuesto por Cohen se denomina *Dummy Variable Adjustment*.

Se observa que estos modelos difieren por la presencia de la variable muda relativa a HOMEPOS, lo cual redundante en la aparición de más coeficientes asociados con las variables (en este caso se añaden los coeficientes $P2=B20=G200$, que no se tenían en el primer modelo).

Al procesar estos modelos con el *software* HLM se tienen los resultados que muestra la tabla A2.9:

Tabla A2.9 Resultados del modelo multinivel incluyendo variable *bandera*

	$Y = P0 + P1*(Variable) + E$					$Y = P0 + P1*(Variable) + P2*(BaVariable) + E$						
Variable	G000	Error	G100	Error	Deviance	G000	Error	G100	Error	G200	Error	Deviance
HOMEPOS	428.7	3.406	7.180	0.456	336599.6	428.8	3.392	7.180	0.456	-47.02	9.13	336558.1

Se observa que los coeficientes G000 y G100 son prácticamente iguales, al igual que los errores estándar de cada uno de ellos. Se aprecia que el nuevo coeficiente G200 puede ser de importancia para la explicación que aporta la variable muda. En cuanto a la calidad del ajuste, se observa que ambos modelos son muy similares, lo cual aporta una evidencia en favor de la pertinencia de la imputación propuesta.

La tabla A2.10 presenta el análisis realizado con todas las variables que tuvieron alguna imputación, siguiendo los dos modelos indicados.

Tabla A2.10 Diferencia entre modelos con variable bandera y sin ella

Primer nivel												
Y = P0 + P1*(Variable) + E						Y = P0 + P1*(Variable) + P2*(BaVariable) + E						
Variable	G000	Error	G010	Error	Deviance	G000	Error	G100	Error	G200	Error	Deviance
HOMEPOS	428.7	3.406	7.18	0.456	336599.6	428.8	3.392	7.18	0.456	-47.02	9.130	336558.1
ESCS	428.8	3.210	9.06	0.335	336268.5	428.9	3.282	9.06	0.193	-39.30	9.595	336233.7
HOMEPOS	428.7	3.406	7.18	0.456	336599.6	428.8	3.392	7.18	0.456	-47.02	9.130	336558.1
ENVOPT	417.9	3.660	-14.93	0.314	334786.1	418.1	3.691	-14.93	0.396	-41.63	5.639	334715.6
HEDRES	425.9	3.492	7.59	0.338	336476.4	426.2	3.399	7.60	0.455	-49.41	3.246	336343.8
WEALTH	429.5	3.330	5.86	0.529	336767.3	429.6	3.313	5.86	0.530	-42.48	9.037	336722.0
ENVAWARE	429.6	3.816	19.46	0.702	334491.4	429.8	3.806	19.46	0.701	-40.03	6.455	334430.1
RESPDEV	414.5	3.800	15.15	0.660	335452.7	414.8	3.753	15.14	0.659	-44.05	2.904	335342.3
SCIEEFF	420.0	3.709	12.94	0.6981	335931.8	420.1	3.678	12.94	0.698	-49.86	6.595	335865.5
ENVPERC	415.7	4.044	8.85	0.6048	336571.4	416.0	4.010	8.85	0.603	-41.84	3.531	336487.3
CULTPOSS	423.3	3.812	6.08	0.4496	336762.2	424.2	3.658	6.08	0.449	-35.07	2.671	336518.8
GENSCIE	417.6	3.706	9.23	0.5380	336346.1	417.8	3.662	9.23	0.539	-40.26	7.317	336277.3
SCINVEST	428.5	3.565	-9.46	0.5607	336343.4	429.5	3.770	-9.45	0.374	-36.45	2.255	336083.2
HISEI	400.0	3.618	0.49	0.0144	336432.8	400.5	3.669	0.50	0.021	-22.86	1.941	336294.6
HISCED	409.0	3.713	3.65	0.1377	336596.7	409.1	3.707	3.65	0.137	-23.19	10.081	336577.0
PARED	400.8	3.553	1.87	0.0697	336584.7	400.9	3.729	1.87	0.094	-23.13	5.220	336565.0
CARINFO	421.0	3.942	-1.02	0.3097	336960.1	421.3	3.907	-1.01	0.309	-46.13	4.259	336843.1
CARPREP	423.0	3.795	-3.23	0.2396	336903.1	423.4	3.868	-3.23	0.391	-52.75	4.310	336753.7
JOYSCIE	414.7	3.731	11.07	1.0209	336459.1	414.8	3.714	11.07	1.022	-45.49	5.213	336401.6
PERSCIE	419.0	3.740	3.54	0.5565	336902.5	419.3	3.690	3.54	0.558	-42.81	4.207	336814.6
SCAPPLY	421.6	3.767	-0.40	0.5243	336969.7	422.6	3.698	-0.40	0.524	-36.54	1.250	336711.7
SCHANDS	423.6	3.669	-4.45	0.4681	336822.3	424.6	3.599	-4.45	0.467	-37.36	1.256	336571.7
SCIEFUT	420.6	3.589	1.58	0.9012	336958.7	420.9	3.535	1.58	0.902	-51.00	4.394	336816.1
SCSCIE	417.2	3.582	8.05	0.7369	336615.9	418.5	3.956	8.13	0.450	-34.30	1.879	336319.0

Segundo nivel												
y= PO+ E donde PO= B00+ B01(Variable)+R0						Y = PO+E donde PO= B00+B01(Variable)+B02(Bavariabale)+R0						
Variable	G000	Error	G010	Error	Deviance	G000	Error	G010	Error	G020	Error	Deviance
SCIPROM	419.3	3.498	16.9	2.096	336837.4	419.9	3.562	16.6	2.265	-17.1	10.192	336832.3
SCMATEDU	432.2	3.922	13.8	1.230	336852.2	431.9	3.928	13.8	1.229	15.3	12.114	336850.2
CLRSIZE	377.2	7.866	1.1	0.131	336896.6	379.2	6.154	1.1	0.122	-15.9	5.570	336888.4
ENVLEARN	419.9	3.626	12.3	2.348	336892.8	420.6	3.766	12.1	2.438	-21.3	13.213	336886.1
PROPCERT	433.2	2.834	-30.4	6.190	336899.9	429.9	4.123	-30.2	6.008	6.6	3.019	336894.5

IMPORTANCIA DE LAS VARIABLES MUDAS

El interés de introducir las variables mudas (*dummy* o *bandera*) para identificar los casos imputados cuando se tiene una omisión en la base de datos, queda en evidencia al comparar la influencia que se tiene en función de la desviación estándar de la variable predictora, en este caso se trata del ESCS.

La variable donde se tiene el mayor impacto es HISEI, el coeficiente es -0.156, que corresponde con -12.38% de la desviación estándar del ESCS (1.26 unidades).

Variable muda		Coeficiente	% de desviación estándar
BAIRATCOMP	Y_{040}	-0.026	-2.06
BASTRATI	Y_{060}	0.027	2.14
BAHISEI	Y_{200}	-0.156	-12.38
BAHISCED	Y_{400}	0.094	7.46

Para comprender mejor la utilidad de las variables mudas, supóngase que dos personas A y B tienen un HISEI de 40 y un ESCS de 0.5, pero se sabe que la persona A omitió su dato HISEI y por ello fue imputado el valor en la base de datos de PISA, siguiendo la metodología indicada en la sección 2.3. La pregunta que puede formularse es: ¿tiene algún impacto en el ESCS que se haya imputado el valor o no?

Por el hecho de haberse imputado, la variable muda toma el valor de 1 y al aplicarse su coeficiente en el modelo multinivel se tiene que el ESCS de la persona A debería ser menor en -0.156 unidades, es decir: 0.344, en lugar de 0.5 que es el valor de la persona B que no fue imputada en el HISEI. Esta diferencia es importante, porque representa más de 12% de la desviación estándar del ESCS, pero esta reducción equivale a que el valor correcto es 32% inferior al valor teórico 0.5.

Como puede verse, al ignorar la influencia de haber imputado un dato se puede cometer un error nada despreciable.

El modelo multinivel puede ser convergente o divergente dependiendo de la escala de los datos y de la presencia de variables *bandera*. Supóngase que el modelo incluye algunas variables cuyo patrón de omisión es similar; entonces las variables *bandera* van a estar altamente correlacionadas incidiendo en que el modelo sea divergente. Se adoptó el siguiente esquema de trabajo para el desarrollo del estudio:

1. En todos los modelos se considera invariablemente la participación de la variable muda en cada caso que haya una imputación.
2. Solo se debe incluir una variable *bandera* si los patrones están altamente correlacionados.
3. Si el modelo es convergente, verificar que el coeficiente de la variable *bandera* no es significativo.
4. Si el modelo no es convergente, eliminar la variable *bandera* para disponer de un modelo con la variable deseada.

A2.7 Respecto de los valores plausibles

La prueba PISA 2006, por tener un diseño de tipo matricial, no pretende aplicar todos los constructos y contenidos a cada estudiante: su objetivo no es informar a cada persona de manera puntual, sino utilizar la prueba para obtener resultados de uso regional o nacional acerca de la población. Cada individuo representa a una cohorte o conjunto de sujetos ante una prueba de bloques incompletos (diseño matricial); pero el resultado es individual, es decir, se trata de un valor discreto de desempeño en el conjunto de ítems que respondió y que deben representar al conjunto de posibles resultados de la agrupación o cohorte de sujetos. Se tiene el problema de que no se desea reportar el resultado puntual de cada estudiante, sino valores representativos de la distribución de puntajes posibles que tendría el grupo o cohorte de sujetos.

Una forma de reportar los valores de la distribución podría ser por medio de un valor medio y de una desviación estándar de la distribución de los sujetos del grupo o cohorte teórica, esta forma no se utiliza en la práctica. Se acostumbra reportar un conjunto de valores representativos de la distribución denominados valores plausibles, que reflejan la gama de posibles valores de la agrupación o cohorte teórica; la media y la desviación estándar de los valores plausibles de todos los sujetos son idénticas a los de la distribución *a posteriori* obtenida de la aplicación de la prueba matricial. La figura A2.4 muestra el puntaje individual dentro de la distribución experimental obtenida en la muestra y la figura A2.5 simula un juego de cinco valores plausibles que representan la distribución de la agrupación o cohorte teórica.

Figura A2.4 Ubicación del puntaje individual

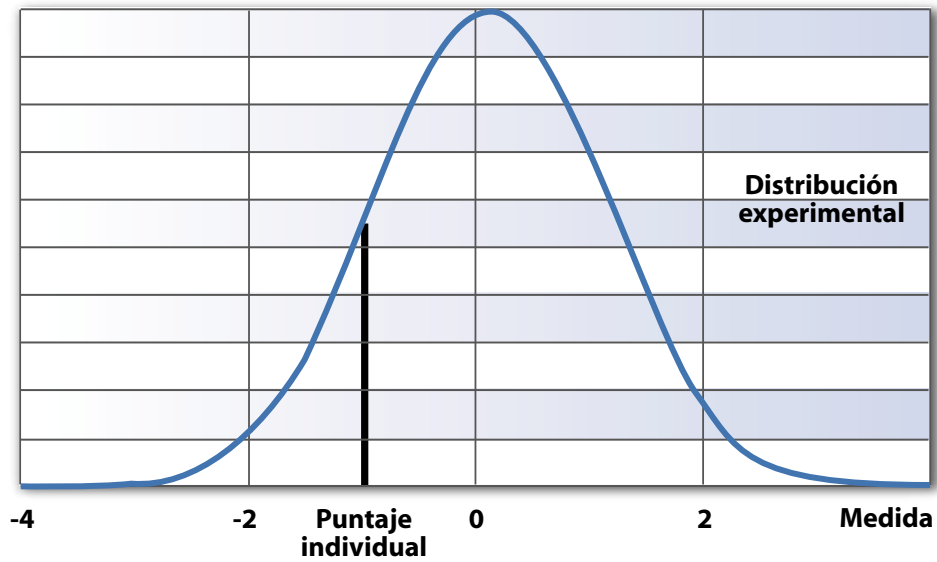
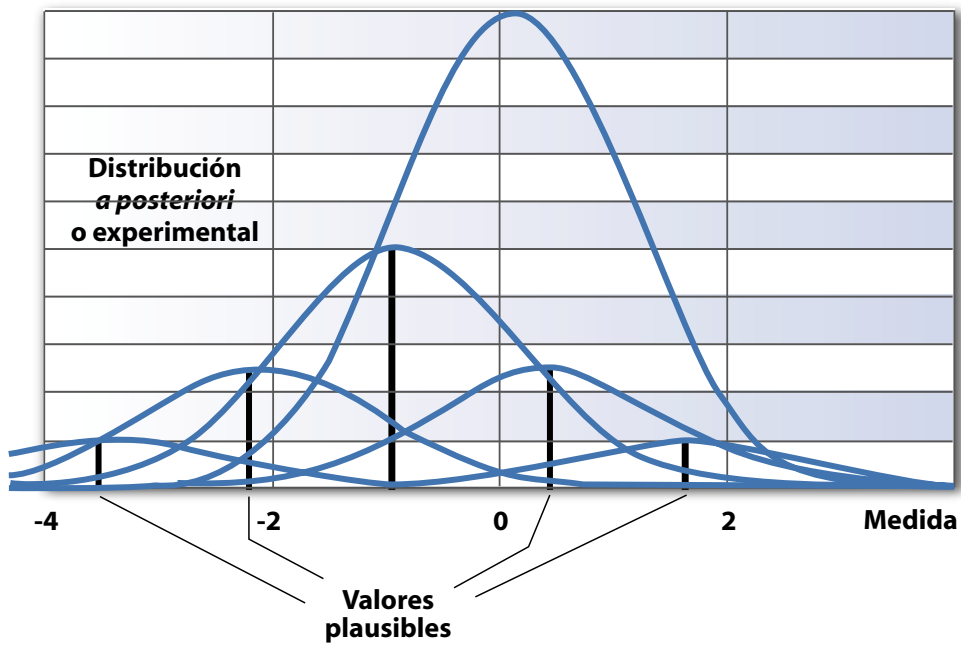


Figura A2.5 Ubicación de los valores plausibles



La cantidad de valores plausibles es arbitraria, se calculan por medio de números aleatorios; los valores se construyen siguiendo una distribución de probabilidad que se deriva de la distribución general de los sujetos y la función de verosimilitud propia del patrón de respuesta de cada sujeto. Siendo números aleatorios no es factible duplicarlos por cualquier persona interesada a menos que se calculen con el mismo *software* utilizado por los responsables del proyecto PISA (donde se define el algoritmo de generación de números aleatorios o pseudo-aleatorios) y con la misma *semilla* de partida. Se acostumbra reportar solamente cinco valores plausibles, que corresponden a igual número de nuevos puntajes posibles que puede tener una persona dentro de la distribución; aunque pueden definirse más valores plausibles si así se desea, se ha visto que cinco es un número razonable para describir la distribución y realizar simulaciones.

Ninguno de los valores plausibles corresponde directamente con la medida de la persona, pero su media es muy similar a la puntuación del sujeto, por ello los cinco valores plausibles de dos personas de la misma medida son diferentes, aunque su media es semejante. Debe observarse también que la media de los valores plausibles se ubica muy cerca o coincide con la media esperada de cada sujeto, pero cada una de las cinco puntuaciones simuladas se ubica de acuerdo con la amplitud total de la distribución, por lo que no se trata de valores dentro del intervalo de confianza de la media en función del error estándar de medida (los cuales serían valores muy cercanos a la media), sino valores con una dispersión similar a la desviación estándar de las puntuaciones observadas. Por ello, en la base de datos de PISA se presentan valores plausibles muy altos y muy bajos, aparentemente lejos de la medida puntual obtenida por el sujeto.

Los valores plausibles se justifican por varias razones técnicas, pero la más importante es porque reducen el sesgo que los estimadores puntuales inducen en los valores poblacionales; el resto de propiedades estadísticas poblacionales se determinan por métodos convencionales, requiriéndose el cálculo de un promedio simple una vez que se tiene los estimados para cada valor plausible. Otra razón importante es porque la variable que se mide en un rasgo latente es continua muy probablemente, pero para cada persona se tiene un resultado puntual, discreto; todas las estimaciones poblacionales serían mejores si partieran de una distribución continua, por ello el uso de los valores plausibles permite subsanar el problema citado, siendo mejor también en el caso de estimaciones poblacionales como pueden ser los percentiles o la mediana.

Tomando en cuenta lo indicado, el proyecto PISA no proporciona la medida puntual que tiene cada sujeto, sino cinco valores plausibles que responden a valores de la distribución continua a posteriori, representando la variabilidad de resultados en la variable latente continua (Adams, R. y Wu, M. L., 2002; Wu, M. L. 2004).

Cuando se realiza un análisis multinivel, se trabaja con los cinco valores plausibles para obtener cinco estimadores (por ejemplo, de la desviación estándar del rendimiento, o del intercepto en un modelo de regresión), los cuales se promedian para proporcionar el valor final de estimación de los parámetros del modelo jerárquico deseado.

El cálculo de los valores plausibles requiere de *software* especializado.

Dado un estudiante de medida de habilidad θ y el vector X del patrón de respuestas a los ítems, se tienen dos propiedades: (1) las medidas de habilidad se distribuyen en forma normal con media μ y desviación estándar σ (descrita por $N(\mu, \sigma^2)$), lo que se denomina modelo poblacional $g(\theta)$ y (2) la probabilidad condicional de respuesta de la persona es $f(X/\theta)$, también denominado modelo de respuesta al ítem. A partir de estos elementos, la función de densidad de probabilidad de la distribución a posteriori de las respuestas del estudiante de medida θ dado el vector X , se obtiene por medio de la expresión:

$$h(\theta/X) = \frac{f(X/\theta)g(\theta)}{\int f(X/\theta)g(\theta)d\theta}$$

Esta función es continua, por lo que difiere del conjunto de valores de respuesta y la medida de un sujeto que son valores discretos. La distribución es de tipo gaussiana con media θ y desviación estándar σ (es decir $N(\theta, \sigma^2)$)

Si se obtienen valores aleatorios respecto de la distribución, como si se tratara de valores observados, se obtiene una distribución empírica *a posteriori* de la función $h(\theta/X)$, que substituye la medida puntual del sujeto. Esta distribución reporta la posible gama de valores que puede adoptar la medida de un estudiante dentro del continuo de la variable latente que mide. El número de valores aleatorios que se pueden generar sobre la distribución empírica *a posteriori* $h(\theta/X)$ no está acotado (puede ser de uno a infinito), pero para fines prácticos se ha optado por determinar cinco valores plausibles que son los que reporta la base de datos de PISA.

Para fines prácticos se cuenta entonces con dos opciones:

- (a) la medida que se debe asignar a un sujeto en particular es la medida puntual de habilidad estimada a partir de X , siguiendo preferentemente el modelo de Rasch
- (b) para estimaciones insesgadas de los parámetros estadísticos poblacionales (media, desviación estándar), análisis de la distribución de medidas (percentiles) y análisis de correlación, resulta conveniente el uso de la distribución $h(\theta/X)$.

A2.8 Criterios para aceptar o rechazar las hipótesis

Ya se comentó en la sección 2.2 que el presente estudio no es experimental, pero se inicia con una pregunta de investigación y su hipótesis de partida en relación con cada dimensión del Modelo de Calidad Educativa del INEE y termina con la construcción del modelo jerárquico que permite corroborar o refutar la hipótesis. Debe resaltarse que no se trata de una hipótesis nula a ser demostrada estadísticamente con los datos provenientes de una muestra diseñada para el estudio.

En lugar de un diseño experimental se cuenta con los datos procedentes de la prueba PISA, con variables provenientes de los cuestionarios de contexto y de otras fuentes disponibles en el INEE, con las cuales se hace un diseño no experimental que requiere establecer los criterios de ajuste del modelo y los criterios de referencia con los cuales se afirma que se tiene una diferencia estadísticamente significativa respecto de una condición planteada desde un principio.

En estudio se utilizan varios criterios: el primero se refiere a la aceptación de la calidad del ajuste del modelo y, una vez aceptado, el segundo tiene que ver con la identificación de las diferencias significativas entre valores de variables.

A2.8.1 Criterios de aceptación del modelo

Dado un modelo se incurre en el problema de estimar qué tanto permite explicar de la varianza de las variables para las hipótesis propuestas. Desde luego se requiere que el modelo sea convergente, pero también se necesita una referencia para contrastar el ajuste del modelo, es por ello que se emplea como referencia para todos los análisis el modelo incondicional de tres niveles.

Para el ajuste del modelo

Cada vez que se hace un análisis con HLM se obtiene un valor relacionado con el logaritmo de la función de verosimilitud ($-2 \times \text{loglikelihood}$) y que se reporta como desajuste (*deviance*). Generalmente este valor es muy grande (varios miles en escala logarítmica) y esto complica la comparación entre modelos, ya que un modelo puede producir un cambio de tan solo una fracción porcentual respecto del valor de origen. Para hacer la comparación de un modelo nuevo contra el modelo incondicional se tienen dos opciones:

1. Reportar la diferencia porcentual entre los dos modelos.

$$D1 = \frac{\text{Deviance}_{\text{nulo}} - \text{Deviance}_{\text{modelo}}}{\text{Deviance}_{\text{nulo}}} \times 100 \quad v[1]$$

No se tiene un valor de referencia, pero se ha observado que diferencias tan pequeñas como 0.2% ya marcan una importante.

2. Bryk, Raudenbush y Congdon (2005) sugieren obtener la diferencia entre los parámetros de desajuste (*deviance*) de dos modelos; ésta tiene una distribución

χ^2 cuyos grados de libertad se obtienen por la diferencia del número de parámetros de ambos modelos. Este esquema permite calcular la significancia en función de un nivel elegido (generalmente del orden de 5%).

$$p(\text{Deviance}_{\text{nulo}} - \text{Deviance}_{\text{modelo}}, \chi^2, \text{gdl}) \leq \alpha$$

$$\text{gdl} = \text{parámetros}_{\text{modelo}} - \text{parámetros}_{\text{nulo}} \quad [2]$$

Se espera que el valor de desajuste debe ser positivo, pero puede ocurrir bajo ciertas circunstancias que el *software* reporte un valor de *deviance* negativo, que implica un buen ajuste pero que el programa no pudo realizar el cálculo para estandarizar el resultado²², debido a la calidad de los datos disponibles, a alguna colinealidad entre variables, a la presencia de las variables mudas o debido a otras causas de interés técnico de los algoritmos de cálculo, pero de poca implicación práctica.

Para la significancia de la diferencia entre valores extremos

Una vez verificado el modelo se puede proceder a dictaminar si hay una diferencia significativa de los valores extremos o entre dos grupos elegidos. Para ello se establecen también dos posibles criterios:

Criterio 1) Establecer una prueba de hipótesis sobre cada coeficiente, donde la hipótesis nula es que el coeficiente no difiere significativamente de cero. Esta prueba de hipótesis es realizada directamente por el *software* HLM, proporcionando el valor del estadístico *t* para juzgar la nulidad del coeficiente ($t = \text{coeficiente elegido} / \text{error estándar}$). Con esta información se pueden tomar decisiones sobre la pertinencia de la variable para participar en el modelado del comportamiento deseado.

Este criterio es muy utilizado al realizar comparaciones entre grupos poblacionales, entidades federativas u otros conjuntos de personas y ha mostrado su utilidad en otros trabajos publicados por el INEE (por ejemplo: Backhoff y col. 2007, a y b). Un inconveniente que se tiene con este enfoque es que sólo puede deducirse el intervalo de confianza de la variable y emitirse un dictamen general que no permite establecer los puntos de corte para definir dónde se tienen cambios cualitativos. De ahí la necesidad de establecer otro modelo más apropiado para los fines que se persiguen en este estudio.

Criterio 2) Establecer un modelo de decisión en términos de diferencias entre subgrupos de estudiantes definidos por rangos de estatus socioeconómico. En este caso se propone el uso del coeficiente *d* de Cohen (1988), comúnmente empleado para analizar la diferencia entre medias tomando en cuenta el efecto de tamaño y también para el análisis del funcionamiento diferencial de ítems (DIF)²³:

$$d = \frac{\mu_a - \mu_b}{\sigma} \quad [3]$$

²² Puede consultarse la sección: *Log likelihood values and negative deviances*, de la guía de ayuda del *software* HLM.

²³ El modelo empleado para DIF puede tomarse como un indicador de equidad, como explican Pardo y col. (2005), Backhoff y Díaz (2005). Es común hacer los estudios de DIF para dictaminar si existe inequidad de género, pero el criterio es aplicable a otras agrupaciones poblacionales.

Donde

$\mu_a > \mu_b$ son las medias de los subgrupos a contrastar y σ es la desviación estándar de la población o de la muestra total.

El criterio de Cohen toma dos valores (en principio medias) que comparten una misma desviación estándar; con ellos calcula la diferencia entre los dos valores estandarizados y la compara contra las áreas común y no común que tendrían dos distribuciones gaussianas cuyas medias corresponden con los valores propuestos. De acuerdo con este criterio la diferencia se puede catalogar como *muy pequeña* si es menor que 0.2, *pequeña* si está entre 0.2 y 0.5; *mediana* de 0.5 a 0.8 y *grande* si es mayor que 0.8. Normalmente son de interés las diferencias *medianas* y *grandes* por ser significativas, en tanto que los casos de diferencias *muy pequeñas* no son significativos. Queda a elección del analista tomar en cuenta o revisar los casos de diferencias *pequeñas*. Para este estudio se consideró importante reportar las diferencias *pequeñas* (por ejemplo de inequidad entre dos niveles socioeconómicos ESCS_i y ESCS_j), entonces d debe sobrepasar 20% de la desviación estándar de los puntajes en Ciencias que, en este caso, es de 15.139 puntos ($0.2 \times \sigma = 0.2 \times 75.696$).

En la escala PISA, el límite para una diferencia pequeña ocurre si:

$$d_{\text{pequeña}} = \frac{\mu_a - \mu_b}{\sigma} = 0.2 \tag{4}$$

La expresión [4] se puede normalizar en relación con el valor esperado para una diferencia pequeña, que se denota como d^* , con lo cual se considera que la diferencia de dos valores $\mu_a > \mu_b$ está en el límite (por ejemplo de inequidad) si:

$$d^* = \frac{\mu_a - \mu_b}{0.2\sigma} = 1 \tag{5}$$

El uso de d^* tiene varias ventajas: no se ve afectado por el tamaño de muestra, satisface el requerimiento de APA de evitar o reducir el uso de hipótesis nulas y permite establecer puntos de corte estableciendo una forma muy simple para dictaminar diferencias entre grupos o valores en las variables: por debajo de 1 no se tienen diferencia significativas y por arriba de 1 las diferencias sí son significativas (tabla A2.11).

Tabla A2.11 Clasificación de la significancia en los valores de d^*

Valor de d^*	Diferencia	Clasificación
Menor que 1	Muy pequeña	No significativa
Mayor que 1 y Menor que 2.5	Pequeña	Significativa
Mayor que 2.5 y Menor que 4	Mediana	Significativa
Mayor que 4	Grande	Significativa

Como ejemplo de uso de este criterio, se presenta la tabla A2.12 tomada del análisis de variables de situación.



Tabla A2.12 Ejemplo de uso de d^*

Variable explicativa	Coeficiente	Categoría para resultado inferior		Categoría resultado superior		Diferencia
NIVEL	44.97	Secundaria = 0	0	Educación media superior = 1	44.97	44.97

Se toman como valores de contraste los límites superior e inferior de cada variable, por ejemplo para el caso de NIVEL los extremos se tienen con Secundaria=0 y Educación media superior=1. Ambos valores se multiplican por el coeficiente obtenido en el análisis multinivel, en este caso (44.97×0) y (44.97×1) por lo que se obtiene una diferencia de 44.97 puntos en la escala de Ciencias de PISA, que al introducirla en la expresión [5], conduce a:

$$d^* = \frac{\mu_a - \mu_b}{0.2\sigma} = \frac{44.97 - 0}{0.2 \times 75.696} = 2.97 > 1$$

Con lo cual se concluye que la diferencia que se obtiene por el nivel educativo es significativo en la categoría de *mediana*.

A2.9 Programas utilizados en el estudio

Un estudio como el que se presenta es factible de ser realizado solamente si se cuenta con una batería de programas informáticos apropiados para cada una de las etapas. Además de las conocidas herramientas de escritorio (en particular la hoja de cálculo), se utilizaron estos programas:

Winsteps, para la recalibración de los ítems y de la prueba por medio del modelo de Rasch.

SPSS empleado como programa de intercambio de archivos para las bases del proyecto PISA 2006, así como para producir las bases empleadas finalmente en los programas de análisis multivariable. También se empleó para el análisis factorial exploratorio sobre las bases de datos en los tres niveles de anidamiento.

AMOS 5 se escogió para hacer los análisis factoriales confirmatorios por medio de ecuaciones estructurales. El *software* admite los archivos procedentes de SPSS y contiene una gran cantidad de herramientas para la formulación gráfica de los modelos, similares a los que se presentaron aquí.

HLM 6 es el programa que se utilizó para hacer el análisis jerárquico lineal de tres niveles de anidamiento. El *software* contiene herramientas para la formulación de las ecuaciones que modelan a la variable dependiente en función de las variables explicativas deseadas, además de herramientas gráficas para la presentación de resultados.

Dentro de las ventajas que tiene HLM 6 se debe contar la posibilidad de realizar modelos con datos faltantes, uso de imputaciones múltiples y cálculo del modelo con valores plausibles y ponderadores para las variables explicativas.

HLM 6 maneja apropiadamente los valores plausibles, como ya se indicó anteriormente. En este caso se parte de un modelo deseado, definido en los tres niveles de anidamiento, el cual contiene una variable dependiente que corresponde con el

primer valor plausible del puntaje de Ciencias. Una vez hecha esa definición, se indica al *software* cuáles son las variables que corresponden a valores plausibles, en este caso del puntaje de Ciencias, con lo cual, de manera automática, el programa se encarga de hacer los cinco modelos de regresión multinivel y obtiene el resultado final como promedio de los cinco modelos realizados. Como puede verse, es un análisis ventajoso porque evita el penoso trabajo de tomar los resultados procedentes de cada corrida para hacer su promedio de forma externa.

A2.10 Archivos de datos

Para el desarrollo del estudio se prepararon tres bases de datos²⁴, dos de ellas (estudiantes y escuelas) proceden de los datos del proyecto PISA en forma directa o mediante algún cálculo para obtener la media por escuela del ESCS por ejemplo. La tercera base de datos (regiones) se construyó a partir de la información disponible de los trabajos realizados previamente por el INEE o alguna otra fuente nacional.

Los campos que sirven de llave o referencia para estas bases son:

REGION, número de identificación de la región (nivel 3)

EsclID, número de identificación de la combinación región-escuela (nivel 2)

EstlID, número de identificación de la combinación región-escuela-estudiante (nivel 1)

Obsérvese que no se utiliza directamente el número de escuela o el número de estudiante, sino en combinación con otro dato (región, escuela), esto es una exigencia del *software* para HLM, que requiere que los datos se encuentren ordenados por medio de los campos llave, debido a que se pueden repetir algunos números, por lo que es más apropiado asignar un número de identificación que combine otros elementos.

Las bases incluyen las columnas con las variables mudas citadas en la sección de variables. En particular, las variables que inician con el prefijo *Ba* hacen referencia a la variable muda de tipo *bandera* que toma un valor de 1 cuando el dato fue imputado, de tal modo que un coeficiente asociado con esta *bandera* permite dimensionar o ajustar la predicción para los casos de omisión, es decir, qué tanto difiere el dato verdadero respecto del dato imputado. Posteriormente, se volverá a revisar el uso de esta *bandera* con los resultados procedentes del análisis multinivel.

Debe hacerse notar en este punto que los datos procedentes de los reportes del INEE, con los cuales se construyó la base de Regiones, para el tercer nivel de animamiento, son muy buenos y permiten completar los datos proporcionados por los cuestionarios de contexto de PISA y hasta hacer comparaciones con los resultados estatales de EXCALE en Lenguaje y Matemáticas.

En estas bases de datos se tiene un total de 30 mil 350 estudiantes, un mil 088 registros de escuelas, 32 entidades federativas y 7 regiones.

²⁴ Los archivos están disponibles en formato de SPSS con estos nombres: EstudianteMexico.SAV, EscuelasMexico.SAV, RegionesMexico.SAV

Anexo 3. Análisis multinivel de los modelos incondicionales

La metodología de análisis de los datos del proyecto PISA, seguida en este estudio, emplea los cinco valores plausibles de la puntuación de Ciencias y la ponderación regional de los datos de la muestra en la población analizada. Para mostrar el efecto que tiene el uso de estos datos, se presenta un ejemplo sobre el cual se establecen cuatro diferentes combinaciones de datos, utilizando modelos incondicionales de tres niveles con la base de resultados de PISA:

1. Primer análisis. Se construye el modelo incondicional con uno solo de los valores plausibles sin ponderación.
2. Segundo análisis. Se repite el modelo incondicional con los cinco valores plausibles sin ponderación.
3. Tercer análisis. El modelo de cinco valores plausibles se enriquece con la ponderación nacional de cada individuo en la muestra de PISA.
4. Cuarto análisis. Es el modelo más completo que utiliza los cinco valores plausibles y la ponderación regional para cada individuo de la muestra de PISA.

En los resultados aquí presentados, se aprecia el impacto de estos cuatro casos, pudiéndose observar que hay algunos cambios en los coeficientes o en el error estándar. En el modelo incondicional la significancia de los valores fijos y de los aleatorios, no parece modificarse, pero este comportamiento puede diferir en los modelos que incluyen variables explicativas. El cuarto modelo es que se utilizó para los análisis presentados en este estudio.

A3.1 Primer caso de modelo incondicional

Modelo incondicional donde la variable dependiente es el primer valor plausible del desempeño en Ciencias sin ponderación.

Tabla A3.1 Resultados del primer modelo incondicional

Efecto fijo			Coeficiente	Error estándar	t	Grados de libertad	P-valor
Puntuación media global en Ciencias		γ_{000}	421.452	3.912	107.725	6	0

Efecto aleatorio			Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
Nivel	Unidad							
1	Estudiante	e_{ijk}	59.222	3507.215				61.0
2	Escuela	r_{ojk}	46.379	2150.966	1081	19366.993	0.000	37.4
3	Región	u_{00k}	9.482	89.905	6	49.590	0.000	1.6

Desajuste	336970.99
-----------	-----------

El intercepto del modelo representa la media global en Ciencias, en este caso 421.45 puntos de la escala PISA. A partir de esta media global, los puntajes de los estudiantes se distribuyen en forma normal, con una desviación estándar de 59.22 puntos, en tanto que la media de las escuelas tiene una desviación estándar ligeramente más pequeña de 46.38 puntos. La dispersión de los resultados regionales tienen la desviación estándar más pequeña de 9.48 puntos.

Los resultados muestran que el nivel I participa con 61% de la varianza, en contraste con 37.4% debido a la varianza de las escuelas y tan solo 1.6% de la varianza es por participación de las regiones. Siendo tan grande la participación de la varianza al nivel de los estudiantes, se vuelve primordial identificar variables explicativas en este nivel.

El valor *t* es el cociente del coeficiente de la variable dividido entre el error estándar que, para ser significativamente distinto de cero, debe tener un valor superior aproximadamente a 2. En la tabla se proporciona el valor de *p* que debe ser menor a 0.05 (5%) para no estar en posibilidad de rechazar la hipótesis nula.

El programa reporta el desajuste o razón de verosimilitud (*deviance*) para el modelo ajustado de acuerdo con las variables propuestas (*deviance* = -2 × *loglikelihood* total, función del logaritmo natural del valor de la función de verosimilitud). Si el modelo se ajustara de forma perfecta a los datos, entonces el desajuste matemáticamente obtenido con el modelo de máxima verosimilitud sería nulo. El parámetro que reporta el programa como desajuste (*deviance*) tiene un valor de 336970.99, lo cual indica que hay una gran dispersión de los datos observados y que deberían agregarse variables explicativas para reducir este desajuste.

A3.2 Segundo caso de modelo incondicional

Para determinar la diferencia con el modelo anterior, se construye el mismo modelo incondicional, pero con todos los valores plausibles sin ponderación.

Tabla A3.2 Resultados del segundo modelo incondicional

Efecto fijo			Coeficiente	Error estándar	t	Grados de libertad	P-valor
Puntuación media global en Ciencias		γ_{000}	421.681	3.959	106.517	6	0.000

Efecto aleatorio			Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
Nivel	Unidad							
1	Estudiante	e_{ijk}	59.361	3523.676				61.1
2	Escuela	r_{0jk}	46.414	2154.274	1081	19300.099	0.000	37.3
3	Región	u_{00k}	9.592	92.015	6	50.783	0.000	1.6

Desajuste	337109.76
-----------	-----------



El intercepto obtenido con los cinco valores plausibles corresponde con una media de 421.68 puntos de la escala PISA. La desviación estándar a nivel de estudiantes es de 59.36 puntos, en tanto que la desviación estándar para las escuelas es de 46.41 puntos; finalmente para las regiones se tiene una desviación de 9.59 puntos.

La varianza a nivel de los estudiantes es de 61.1%, en tanto que se tiene 37.3% de participación en la varianza a nivel de las escuelas y permanece 1.6% de la varianza por parte de las regiones.

El empleo de los cinco valores plausibles mejora el resultado del modelo, porque el desajuste es de 337109.76, lo cual representa una diferencia de 0.057% respecto al desajustes del modelo anterior.

A3.3 Tercer caso de modelo incondicional

En este caso se toman en cuenta todos los valores plausibles con la ponderación estandarizada de acuerdo con las proporciones en la muestra original.

Tabla A3.3 Resultados del tercer modelo incondicional

Efecto fijo		Coeficiente	Error estándar	t	Grados de libertad	P-valor
Puntuación media global en Ciencias	γ_{000}	418.035	10.633	39.314	6	0.000

Efecto aleatorio			Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
Nivel	Unidad							
1	Estudiante	e_{ijk}	51.958	2699.647				48.7
2	Escuela	r_{0jk}	49.171	2417.798	1081	31562.119	0	43.6
3	Región	u_{00k}	20.741	430.209	6	303.595	0	7.8

Desajuste (Promedio)	338978.20
----------------------	-----------

No.	Nivel		Componente de varianza en el modelo nulo	Componente de varianza en este modelo	Diferencia de varianza %
1	Estudiante	e_{ijk}	3507.22	2699.65	23.0
2	Escuela	r_{0jk}	2150.97	2417.80	-12.4
3	Región	u_{00k}	89.90	430.21	-378.5

Al utilizar los cinco valores plausibles y la ponderación se produce un cambio interesante. Ahora son más cercanas entre sí las participaciones de los niveles 1 y 2 en los efectos aleatorios en la varianza total; la fracción de varianza que podría explicarse en el nivel de estudiantes es de 48.7% y de 43.6% a nivel de escuelas. Las regiones incrementan su participación en la varianza a 7.8%, lo cual es notablemente más importante que el caso sin ponderación. Si se calculan las diferencias respecto del primer modelo nulo, se aprecia el impacto de lo que podría denominarse ganancia en el nivel 3, al tener una diferencia de 378.5%, que indica el mayor interés que se tiene al modelar variables en el nivel de regiones.

En este modelo, el intercepto reporta una media global de 418.04 puntos de la escala de Ciencias de PISA. La desviación estándar a nivel de estudiantes es de 51.96 puntos, en tanto que la desviación estándar para las escuelas es de 49.17 puntos; finalmente para las regiones se tiene una desviación de 20.74 puntos.

El empleo de los cinco valores plausibles implica un mejor resultado del modelo, porque el desajuste es de 332288.70, es decir, hay 1.4% de mejora respecto al primer modelo donde sólo se utilizaban los valores plausibles sin ponderación.

A3.4 Cuarto caso de modelo incondicional

El último modelo incondicional toma en cuenta todos los valores plausibles y la ponderación estandarizada por región.

Tabla A3.4 Resultados del cuarto modelo incondicional

Efecto fijo		Coefficiente	Error estándar	t	Grados de libertad	P-valor
Puntuación media global en Ciencias	γ_{000}	418.169	10.730	38.972	6	0.000

Efecto aleatorio			Desviación estándar	Componente de varianza	Grados de libertad	χ^2	p-valor	Porcentaje de varianza
Nivel	Unidad							
1	Estudiante	e_{ijk}	58.455	3416.982				49.4
2	Escuela	r_{0jk}	55.346	3063.174	1081	24936.461	0	44.3
3	Región	u_{00k}	21.002	441.091	6	198.675	0	6.4

Desajuste (Promedio)	336665.13	← Modelo de referencia
----------------------	-----------	------------------------

No.	Nivel		Componente de varianza en el modelo nulo	Componente de varianza en este modelo	Diferencia de varianza %
1	Estudiante	e_{ijk}	3507.22	3416.98	2.6
2	Escuela	r_{0jk}	2150.97	3063.17	-42.4
3	Región	u_{00k}	89.90	441.09	-390.6

En este modelo se utilizaron los cinco valores plausibles y la ponderación regional, con lo cual se tienen los cambios más importantes. De nuevo se observa que son comparables las participaciones de los efectos aleatorios de los niveles 1 y 2 en la varianza total (49.4 y 44.3% respectivamente); a nivel de las regiones se tiene 6.4% de intervención en la varianza, lo cual es muy cercano al caso anterior ponderado respecto de la muestra nacional.

En este caso, el intercepto que reporta la media global es de 418.17 puntos de la escala de Ciencias de PISA. La desviación estándar a nivel de estudiantes es de 58.46 puntos, en tanto que la desviación estándar para las escuelas es de 55.35 puntos; finalmente para las regiones se tiene una desviación de 21.00 puntos. El empleo de los cinco valores plausibles y la ponderación mejora ligeramente el ajuste,

porque ahora el valor de desajuste es de 332288.70, lo cual representa 0.09% de mejora respecto al primer modelo, donde se utilizaban sólo los valores plausibles sin ponderación.

Este es el último modelo nulo que se presenta como ejemplo para el análisis multi-nivel de los datos del proyecto PISA. Los valores de varianza y de desajuste del cuarto modelo son los que se emplean como referencia para dictaminar la mejora que proporcionan los modelos de los restantes análisis desarrollados sobre las dimensiones del Modelo de Calidad Educativa del INEE.

Índice de tablas

3.1	Resultados del análisis factorial exploratorio del nivel 1 (valores absolutos)	40
3.2	Resultados del análisis factorial exploratorio del nivel 2 (valores absolutos)	41
3.3	Resultados del análisis factorial exploratorio del nivel 3 (valores absolutos)	41
3.4	Resultados del modelo incondicional	44
3.5	Variables explicativas del modelo de la situación del estudiante	45
3.6	Modelo multinivel de las variables que integran la situación del estudiante	47
3.7	Valores de d^* para las variables del modelo situación	50
3.8	Variables del modelo de equidad	53
3.9	Resultados del modelo multinivel del índice socioeconómico ESCS	56
3.10	Resultados del primer modelo de equidad con ESCS	58
3.11	Resultados del modelo de equidad con término cuadrático	59
3.12	Valores de d^* para las variables del modelo de equidad con el ESCS	60
3.13	Valores de d^* para las variables del modelo de equidad con el ESCS2	60
3.14	Valores de ESCS para cambios unitarios en d^*	64
3.15	Comparación de puntajes de inequidad entre regiones	66
3.16	Variables que componen el modelo de eficiencia	68
3.17	Resultados del modelo multinivel del índice HOMEPOS	71
3.18	Valores de d^* para las variables del modelo de HOMEPOS	72
3.19	Resultados del modelo multinivel de eficiencia	73
3.20	Valores de d^* para las variables del modelo de eficiencia	74
3.21	Variables del modelo de impacto	77
3.22	Resultados del primer modelo de impacto	80
3.23	Resultados del segundo modelo de impacto	81
3.24	Valores de d^* para las variables que integran el primer modelo de impacto	82
3.25	Valores de d^* para las variables que integran el segundo modelo de impacto	82
3.26	Variables del modelo de eficacia	85
3.27	Resultados del modelo de la dimensión de eficacia con variables de nivel 3	88
3.28	Resultados del modelo de la dimensión de eficacia con variables de los niveles 2 y 3	89
3.29	Valores de d^* para las variables del modelo de eficacia 1	90
3.30	Valores de d^* para las variables del modelo de eficacia 2	90

3.31	Variables potencialmente significativas en el nivel 3	92
3.32	Variables del modelo de pertinencia y relevancia	94
3.33	Resultados del modelo de la dimensión de pertinencia y relevancia	96
3.34	Valores de d^* para las variables del modelo de pertinencia y relevancia	97
3.35	Porcentajes de varianza explicada por los modelos en cada nivel	104
A1.1	Distribución de variables por cada nivel	128
A2.1	Variables nivel estudiante	134
A2.2	Variables nivel escuela	140
A2.3	Variables de la base de datos de regiones (nivel estado-región)	141
A2.4	Diferencias por género de variables no cognitivas de PISA	148
A2.5	Variables del ejemplo de ecuaciones estructurales	160
A2.6	Variables del ejemplo de modelos multinivel	163
A2.7	Términos fijos del modelo multinivel de las variables de situación del estudiante	164
A2.8	Términos aleatorios del modelo multinivel de la situación del estudiante	165
A2.9	Resultados del modelo multinivel incluyendo variable <i>bandera</i>	167
A2.10	Diferencia entre modelos con variable <i>bandera</i> y sin ella	168
A2.11	Clasificación de la significancia en los valores de d^*	176
A2.12	Ejemplo de uso de d^*	177
A3.	Resultados del primer modelo incondicional	179
A3.2	Resultados del segundo modelo incondicional	180
A3.3	Resultados del tercer modelo incondicional	181
A3.4	Resultados del cuarto modelo incondicional	182

Índice de figuras

1.1	Modelo de Calidad Educativa del INEE	21
3.1	Distribución de la dificultad de los ítems de Ciencias en la prueba PISA	38
3.2	Mapa de Wright de la prueba de Ciencias de PISA	39
3.3	Modelo estructural de la situación del estudiante	46
3.4	Diferencias en desempeño de Ciencias por variables de la situación del estudiante	49
3.5	Modelo estructural de la dimensión de equidad en el nivel 1	54
3.6	Modelo estructural de la dimensión de equidad en el nivel 2	55
3.7	Diagrama de dispersión ESCS vs. Puntuación de Ciencias, modelo de primer grado	61
3.8	Comparación del modelo multinivel y el modelo de regresión lineal del ESCS	62
3.9	Diagrama de dispersión ESCS vs. Puntuación de Ciencias, modelo de segundo grado	63
3.10	Intervalos de inequidad a partir del modelo multinivel y del criterio de Cohen	65
3.11	Diferencias por inequidad en las siete regiones	66
3.12	Modelo estructural de la dimensión de eficiencia con variables de nivel 1	69
3.13	Modelo estructural de la dimensión de eficiencia con variables de nivel 2	70
3.14	Valores de d^* para las variables del modelo de eficiencia	74
3.15	Modelo estructural de la dimensión de impacto 1	78
3.16	Modelo estructural de la dimensión de impacto 2	79
3.17	Valores de d^* para las variables de la dimensión de impacto	83
3.18	Modelo estructural de la dimensión de eficacia nivel 2	86
3.19	Modelo estructural de la dimensión de eficacia nivel 3	87
3.20	Distribución de frecuencias de la proporción de docentes certificados en las escuelas	91
3.21	Modelo estructural de la dimensión de pertinencia y relevancia con variables de nivel 1	95
3.22	Modelo estructural de la dimensión de pertinencia y relevancia con variables de nivel 2	96

Figura A2.1	Regionalización socioeconómica propuesta por el INEE	146
Figura A2.2	Regionalización socioeconómica propuesta por el INEE	147
Figura A2.3	Modelo de ejemplo de ecuaciones estructurales	160
Figura A2.4	Ubicación del puntaje individual	171
Figura A2.5	Ubicación de los valores plausibles	171