

INVESTIGACIONES

*PERFIL DE RENDIMIENTO DE CHILE EN LA SUBESCALA DE
REPRESENTACION DE DATOS TIMSS 1999*

Profile of Chilean achievement in the TIMSS 1999 Data Representation Sub-scale

*Prof. María José Ramírez**

International Study Center, Boston College, Lynch School of Education, Beaton 188,
Chestnut Hill, MA 02467, U. S. A.

Resumen

Por primera vez Chile participó en la Repetición del Tercer Estudio Internacional de Matemáticas y Ciencias de IEA (TIMSS 1999), evaluando los conocimientos de una muestra representativa a nivel nacional de 5.907 alumnos de octavo básico, y recogiendo valiosa información de contexto sobre los alumnos, sus profesores y sus colegios. Este artículo analiza los resultados obtenidos en representación de datos, la subescala de matemáticas con los mejores resultados relativos para Chile. Aquí se analiza el efecto de cinco características de las preguntas incluidas en la prueba: formato, habilidad cognitiva, contenido, intención curricular e implementación curricular. Las conclusiones apuntan a que el mejor rendimiento en representación de datos puede explicarse por un fenómeno de aprendizaje informal de las matemáticas combinado con un efecto de formato de las preguntas.

Palabras claves: Rendimiento matemático, resultados de pruebas, octavo básico, educación comparada

Abstract

For the first time, Chile participated in IEA's Trends in Mathematics and Science Study (TIMSS 1999), testing a nationally representative sample of 5907 eighth graders, and collecting vast background information about the students, their teachers and their schools. This article analyzes student's achievement in data representation, the sub-scale that showed the best relative math results for Chile. The effect of five item-related variables was analyzed: format, cognitive skill, sub-content area, curricular intentions, and curricular implementation. Conclusions indicate that the better performance in data representation can be mainly explained because of a "street mathematics" phenomenon combined with an item-format effect.

Key words: Mathematics achievement, test results, grade 8, comparative educación.

* Quisiera agradecer a Laura Misas por su apoyo en el proceso de revisión y edición de este artículo.

INTRODUCCION

En TIMSS 1999 (Tercer Estudio Internacional de Matemáticas y Ciencias), Chile se ubicó en el lugar número 35 entre los 38 países participantes. Su rendimiento promedio fue de 392 puntos¹, y su desempeño comparable a países como Turquía, Jordania, Irán, e Indonesia (Mullis *et al.* 2000, cap 1.). La prueba de matemáticas incluía cinco áreas de contenido: “fracciones y sentido numérico”, “medición”, “representación de datos”, “geometría” y “álgebra”. En todas las áreas el desempeño chileno fue de los más pobres de entre los países participantes. No obstante, se reportan diferencias significativas en el desempeño relativo de estas áreas. Mientras los resultados más pobres fueron los de álgebra y números y fracciones, los mejores fueron alcanzados en el área de representación de datos (Mullis *et al.* 2000, cap. 3).

¿Qué explica este mejor desempeño en representación de datos en comparación con las otras áreas de contenido? ¿Qué se puede aprender de este dominio que ayude a mejorar el desempeño en los otros? En rigor, responder a estas preguntas implica un análisis comparativo entre las distintas áreas evaluadas. Sin embargo, un buen punto de partida es mirar en profundidad en una de las áreas evaluadas: Representación de datos.

Los análisis que aquí se presentan se guían por cinco hipótesis que surgen de la literatura especializada sobre factores que afectan el rendimiento en pruebas estandarizadas:

1. *Hipótesis sobre el formato de pregunta.* ¿Pueden los distintos formatos de preguntas utilizados en la prueba ayudar a explicar el patrón de respuesta observado en la subescala de representación de datos? ¿Cómo varía el rendimiento de los alumnos en función de ese formato? ¿Son las preguntas de desarrollo más difíciles de responder que las preguntas de selección múltiple?
2. *Hipótesis sobre expectativas de rendimiento.* ¿Se puede explicar la gradiente de rendimiento en la subescala de acuerdo a las expectativas de rendimiento de las preguntas (habilidades cognitivas en jerga TIMSS)? ¿Son las preguntas más difíciles las que requieren el uso de habilidades cognitivas superiores? ¿Y las más fáciles sólo demandan el uso de habilidades básicas?
3. *Hipótesis sobre subáreas de contenido.* ¿Es posible explicar el perfil de rendimiento chileno de acuerdo a los distintos subcontenidos evaluados en representación de datos? ¿Tienen un nivel de dificultad similar aquellas preguntas que evalúan un mismo subcontenido?
4. *Hipótesis sobre currículum intencionado.* El currículum intencionado se refiere a las intenciones y objetivos curriculares que especifican qué se espera que aprendan los alumnos en el colegio. El currículum oficial es el documento prototipo de currículum intencionado. De acuerdo a esta definición, la pregunta de investigación deviene: ¿Las preguntas que los alumnos de octavo básico deberían ser capaces de responder –de acuerdo a las especificaciones del currículum– fueron, efectivamente, más fáciles de responder que las preguntas que, de acuerdo al currículum, no son aptas para el nivel?
5. *Hipótesis sobre el currículum implementado.* El currículum implementado se refiere a la instrucción que efectivamente reciben los alumnos en la sala de clases. La

¹ La escala TIMSS tiene un promedio de 500 y una desviación estándar de 100.

pregunta que surge entonces es: ¿Presentan los alumnos que pasaron la materia necesaria para responder las preguntas un mejor desempeño que aquellos que aún no pasaban la materia?

En la siguiente sección se explica el procedimiento seguido para responder a cada una de estas preguntas, para luego presentar los principales resultados del estudio. En el último capítulo se discuten estos resultados desde una perspectiva más amplia.

METODOLOGIA

Fuentes de Información. En Chile, una muestra representativa de 5.907 alumnos de octavo básico, provenientes de 185 colegios, rindió la prueba TIMSS en noviembre de 1998². Los profesores de matemática de estos alumnos respondieron un cuestionario proveyendo información sobre sus antecedentes personales y nivel de preparación para enseñar, temas cubiertos en la sala de clases y prácticas instruccionales, entre otros. La coordinación nacional TIMSS³ entregó información sobre el sistema educacional chileno, el marco curricular y sus énfasis, el sistema nacional de evaluaciones, entre otras muchas informaciones de contexto.

Toda esta información se encuentra en la *Base de Datos Internacional TIMSS 1999* (2001), principal fuente de información de este estudio. Ciertos datos fueron directamente sacados del *Reporte Internacional de Matemáticas TIMSS 1999* (Mullis *et al.* 2000). Los detalles acerca de los procedimientos utilizados para recoger los datos y producir esta información pueden encontrarse en la serie de publicaciones TIMSS 1999 (Mullis *et al.* 2000; Martin, Gregory & Stemler 2000; González & Miles 2001).

Información del Estudio de Análisis Curricular TIMSS también fue utilizada para este informe. Si bien Chile no participó en el estudio original (Schmidt, Raizen, Britton, Bianchi & Wolfe 1997), la misma metodología fue utilizada en un proceso de recolección de datos llevado a cabo en 1999. El método consistía en un enfoque analítico en donde el currículum y los libros de estudio fueron codificados de acuerdo al contenido y las expectativas de rendimiento presentadas en cada “bloque”. Los “bloques” fueron la unidad básica de los documentos analizados, y su tamaño varió desde una simple frase (ej., un objetivo curricular) hasta un párrafo breve (ej., una actividad instruccional). Esta información fue utilizada para dar cuenta de las intenciones curriculares y de las experiencias de matemática que más probablemente habían tenido los alumnos de octavo básico. En Chile, el libro de texto utilizado por el 90% de los estudiantes fue codificado, así como también el currículum vigente el año de aplicación de TIMSS⁴ (Ministerio de Educación, 1999). Ya que este currículum fue reemplazado por uno nuevo en el año escolar que se inició en marzo de 2002, me referiré a él como el currículum antiguo.

² Chile, al igual que los demás países del hemisferio sur, rindió la prueba al final del año escolar de 1998; los países del hemisferio norte recogieron sus datos en 1999.

³ La coordinación nacional TIMSS 1999 fue asumida por el Ministerio de Educación, Programa SIMCE, Unidad de Estudios Internacionales.

⁴ El currículum analizado fue *Planes y Programas de la Educación General Básica* (Ministerio de Educación, 1980). El libro seleccionado fue *Matemática 8* (Lara, Luque & Mendoza 1998).

Procedimiento. Para apreciar mejor el desempeño de Chile en las 21 preguntas de la subescala de representación de datos, estas preguntas fueron numeradas del 1 al 21, de acuerdo al nivel de dificultad presentado para los alumnos chilenos. El nivel de dificultad se basó en el porcentaje promedio de respuestas correctas obtenido en cada ítem (valor p). La pregunta más fácil –la que fue correctamente respondida por la mayor proporción de estudiantes– fue llamada Pregunta 1; el segundo ítem con la mayor proporción de respuestas correctas fue llamado Pregunta 2, y así consecutivamente hasta que la pregunta más difícil recibió el nombre de Pregunta 21. En todos los gráficos presentados en este informe, el número de la pregunta se refiere siempre al mismo ítem.

Los ítemes fueron luego ordenados de menor a mayor dificultad, y representados en un gráfico cuyo eje X presenta las Preguntas 1 a 21, y el eje Y indica el porcentaje de respuestas correctas. El resultado de este procedimiento es un perfil de rendimiento para la subescala de representación de datos, en donde los puntos de intersección forman una suave gradiente que da cuenta del nivel de dificultad observado para el conjunto de ítemes. Este mismo procedimiento fue repetido varias veces, y sólo se cambió la simbología utilizada para representar los puntos, de modo que las figuras pudieran incluir información sobre una tercera variable. Esta tercera variable estaba directamente relacionada con cada una de las preguntas de investigación: formato de las preguntas, expectativas de rendimiento, subáreas de contenido, intenciones curriculares e implementación curricular.

El porcentaje de respuestas correcto para Chile y el promedio internacional fue directamente obtenido de los archivos de datos TIMSS. La información concerniente a los distintos tipos de preguntas –formato, habilidades cognitivas y contenido evaluado e intenciones curriculares– fue obtenida de los archivos de información sobre ítemes que también se encuentran en la base de datos TIMSS.

La clasificación de preguntas según habilidades cognitivas requirió cierta reorganización de las categorías originales. El informe internacional menciona cinco expectativas de rendimiento: “conocer”, “realizar procedimientos rutinarios”, “resolución de problemas e investigación”, “razonamiento matemático” y “comunicación” (Mullis *et al.* 2000: 318). La base de datos TIMSS provee una clasificación más específica consistente en 10 categorías: “reconocer objetos y propiedades matemáticas”, “representar”, “realizar procedimientos rutinarios”, “realizar procedimientos complejos”, “resolver”, “predecir”, “resolver, describir, discutir”, “formular y aclarar problemas y situaciones”, “generalizar” y “reconocer equivalentes”. Todas menos las tres últimas categorías se aplican a la subescala de representación de datos.

Para los propósitos de este informe, el segundo sistema de clasificación fue utilizado, ya que presentaba mayor relación con las preguntas de representación de datos. Sin embargo, se introdujeron algunos cambios con el propósito de dar mayor sentido a dicha clasificación. Específicamente, cinco de las categorías originales fueron reorganizadas en sólo dos. Así, “reconocer objetos y propiedades matemáticas” y “representar” fueron fusionados en un solo grupo; y “resolver”, “predecir” y “resolver, describir, discutir” conformaron un segundo.

VER Para calcular el porcentaje de respuestas correctas obtenidas por los alumnos que ya habían “pasado la materia” y los que no, fue necesario combinar la información de rendimiento de los alumnos y la información del cuestionario de profesores. A éstos se les preguntó cuándo vieron sus alumnos los contenidos evaluados en la prueba. En la

sección de “Representación de datos, análisis y probabilidades”, a los profesores se les preguntó sobre las tres subáreas de contenido que conforman la subescala de representación de datos:

- Representación e interpretación de datos en gráficos y tablas.
- Promedio aritmético⁵.
- Probabilidades simples (comprensión y cálculo).

Para cada uno de estos tópicos los profesores debían responder si el tópico fue: a) enseñado antes de este año; b) enseñado 1 a 5 períodos durante este año; c) enseñado más de 5 períodos durante este año; d) No ha sido enseñado aún; e) No sabe. Estas respuestas fueron reclasificadas en dos categorías: ya enseñado (incluye categorías primera a tercera) y aún no enseñado (incluye la cuarta categoría solamente). Los “No sabe” –correspondientes al 5% de las respuestas– fueron excluidos del análisis.

Para cada subárea de contenido los alumnos fueron clasificados de acuerdo a las respuestas de sus profesores. Por ejemplo, un alumno puede aparecer como “ya enseñado” en representación e interpretación de datos y en promedio aritmético, pero estar en “aún no enseñado” en probabilidades. En este caso, el alumno es considerado como “ya enseñado” en las dos primeras subáreas y como “no enseñado aún” en probabilidades.

RESULTADOS

Perfil de Rendimiento de Chile versus Perfil de Rendimiento Internacional. La figura 1 presenta el nivel de dificultad para Chile (porcentaje de respuestas correctas) en los 21 ítems de la subescala de representación de datos. Como se explicó previamente, los ítems se ordenaron desde el más fácil (Pregunta 1) al más difícil (Pregunta 21), de modo tal que formaran una gradiente de rendimiento para el país. El gráfico también muestra el promedio de dificultad internacional⁶ para cada pregunta. El promedio internacional se incluyó con el objeto de contar con un referente externo que permitiera juzgar el desempeño nacional.

Este gráfico merece varios comentarios. En primer lugar, es notorio el amplio rango de niveles de dificultad observado en la subescala de representación de datos. Mientras 71,60% de los alumnos respondió correctamente la Pregunta 1, sólo 8,15% obtuvo la respuesta correcta en la Pregunta 21⁷. Si se comparan las diferencias de rendimiento de

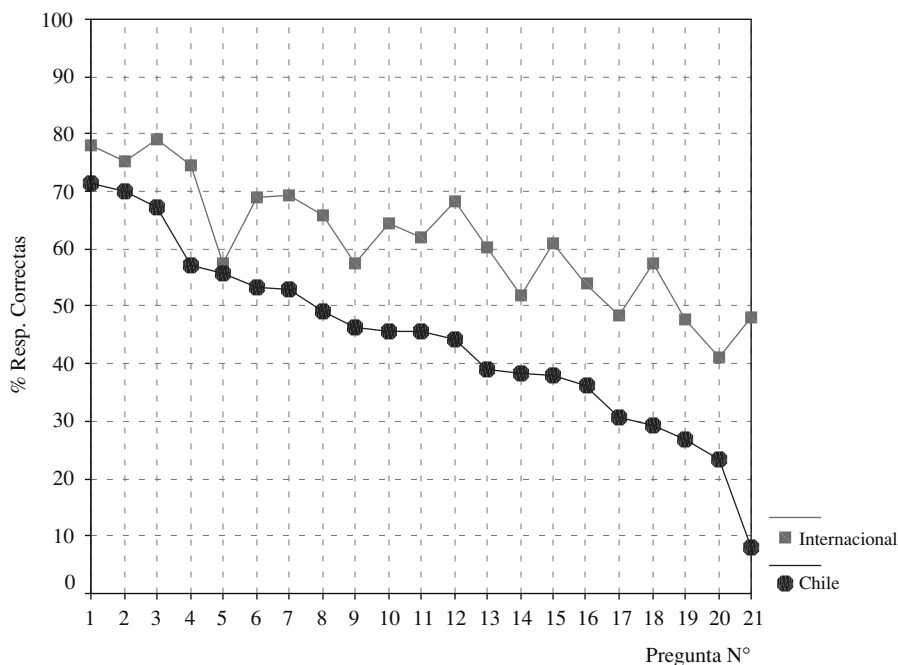
⁵ A los profesores se les preguntó por el promedio aritmético y no por rangos y promedios –el nombre oficial de esta categoría. Es muy probable que este cambio obedezca a lo que realmente mide la Pregunta 2 (único ítem perteneciente a esta subárea de contenido), que sólo pide comparar dos promedios.

⁶ El nivel de dificultad internacional se calculó tomando en consideración el promedio nacional de los 38 países participantes en el estudio.

⁷ El informe internacional (Mullis *et al.* 2000: 65) dice que el porcentaje de respuestas correctas de Chile en la pregunta 21 fue de 5% y no de 8,15%. La diferencia se explica por el hecho de que en el informe internacional sólo se cuentan los alumnos que obtuvieron puntaje máximo (2 puntos), mientras que en este reporte también se consideran a aquellos alumnos que obtuvieron crédito parcial (1 punto). En el cálculo se asignó a los segundos la mitad del peso que se les asignó a los primeros con el objeto de dar cuenta de estas diferencias de puntaje.

Figura 1

Perfil de Rendimiento Nacional e Internacional



Fuente: Base de datos TIMSS 1999.

Chile y las diferencias de rendimiento para el promedio internacional, tenemos que las primeras son una y media vez más grandes que las segundas –63,45 versus 38,20 puntos porcentuales de diferencia entre la pregunta más fácil y la más difícil para Chile y el promedio internacional, respectivamente.

Segundo, y directamente relacionado al párrafo anterior, si se comparan las diferencias de nivel de dificultad para Chile y el promedio internacional, se observa que estas diferencias crecen a medida que nos movemos hacia los ítemes más difíciles. Así, mientras en la Pregunta 1 ambas líneas están relativamente cerca una de la otra (reflejando una diferencia de 6,60 puntos entre Chile y el promedio internacional), en la Pregunta 21 las líneas están muchísimo más separadas (dando cuenta de una diferencia de 40,10 puntos). Una clara excepción a este patrón es la Pregunta 5, que presenta casi el mismo nivel de dificultad para Chile y el promedio internacional –con 1,70 puntos a favor de este último.

¿Sucede algo especial en la Pregunta 5 que explique su comportamiento diferencial? Esta es una pregunta de *selección múltiple* que indaga sobre la *probabilidad* de obtener un cierto resultado después de lanzar por quinta vez una moneda al aire. El currículum antiguo no consideraba al área de probabilidades como un contenido a ser enseñado a los alumnos de octavo básico, y sin embargo 55,70% de los alumnos chilenos respondieron

correctamente esta pregunta. De hecho, este ítem resultó ser uno de los más fáciles de la subescala de representación de datos.

La Pregunta 5 exigía *recordar objetos y propiedades matemáticas*, es decir, formaba parte de las preguntas con menor demanda cognitiva de acuerdo a la clasificación de expectativas de rendimiento de la prueba. Una hipótesis es que el desempeño nacional igualó al internacional por el bajo nivel de demanda cognitiva de la pregunta. Sin embargo, esta explicación no es satisfactoria, ya que el ítem sería más fácil no sólo para los chilenos, sino que para todos los alumnos que rindieron la prueba en otros países del mundo. Una hipótesis más plausible se refiere a la familiaridad del problema presentado a los alumnos. Que la probabilidad de obtener cara o sello al lanzar una moneda es siempre la misma –sin importar cuál fue el resultado del lanzamiento anterior– es algo que se puede aprender informalmente, como parte de las “matemáticas de la calle”. Esta línea de argumentación será desarrollada a lo largo de este trabajo.

Un tercer comentario: en términos generales, se puede afirmar que las preguntas más difíciles para Chile también fueron las preguntas más difíciles en términos internacionales, y viceversa. Lo que cambia drásticamente entre un grupo y otro es el rendimiento promedio obtenido en la subescala de representación de datos. Esto se observa en la distancia que separa ambas líneas de rendimiento en el gráfico. En promedio, la diferencia es de 17,25 puntos entre ambas líneas, una diferencia importante y significativa ($p < 0,005$). Esta distancia también se puede describir usando la escala de puntajes TIMSS: mientras Chile obtuvo 429 puntos en la subescala de representación de datos, el promedio internacional en esta escala fue de 487 puntos (Mullis *et al.* 2000: 97).

El cuarto y último comentario se refiere a la Pregunta 21. Este ítem definitivamente se sale de la gradiente descrita por el patrón de rendimiento nacional. Al pasar de una pregunta a otra, las diferencias promedio son de dos o tres puntos porcentuales. Sin embargo, la Pregunta 21 presenta 15% menos de respuestas correctas que la pregunta que la antecede en nivel de dificultad. El perfil internacional no sigue este patrón. Un cambio brusco en el valor p también se observa al pasar de la Pregunta 3 a la 4 (con 10 puntos de diferencia). Pero en este caso las diferencias dan cuenta de un patrón internacional seguido por todos los países. Como se indica en la siguiente sección, una posible explicación para estas observaciones es el formato especial de la Pregunta 21.

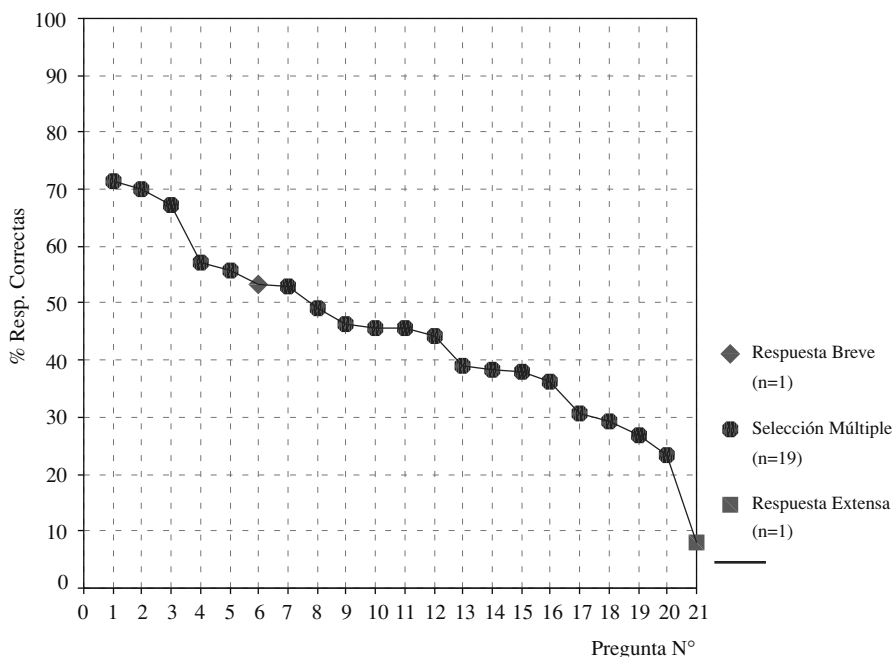
Hipótesis sobre el formato de las preguntas. La prueba TIMSS 1999 incluía tres tipos de formato de preguntas (Martin, Gregory & Stemler 2000, cap. 3):

- *Ítems de selección múltiple.* Los alumnos deben seleccionar una respuesta correcta de entre 4 ó 5 opciones. Estos ítems valían un punto, y constituían tres cuartos de las preguntas de la prueba de matemática (125 de 162 preguntas).
- *Ítems de respuesta breve.* Los alumnos deben generar sus propias respuestas, escribiendo una operación, un número, o algún tipo de respuesta breve. Estos ítems conformaban algo más del 10% de las preguntas de la prueba (21 de 162 preguntas) y valían un punto.
- *Ítems de respuesta extensa.* Los alumnos deben desarrollar una respuesta más larga y elaborada en apoyo a sus respuestas, escribiendo frases completas o respondiendo ítems con dos o tres secciones. Estos ítems constituían algo menos del 10% de las preguntas (16 de 162 preguntas) y valían dos o tres puntos.

Con el objeto de probar si hay un *efecto del formato de las preguntas* que pudiera ayudar a explicar el perfil de rendimiento de Chile, se procedió a graficar de nuevo los 21 ítems de la subescala de representación de datos. Esta vez se usaron diferentes marcas para indicar los distintos tipos de formato (figura 2).

Figura 2

Perfil de Rendimiento según Formato de las Preguntas



Fuente: Base de datos TIMSS 1999.

La figura muestra que 19 de las 21 preguntas que conforman la subescala de representación de datos corresponden a preguntas de selección múltiple. Esta es una proporción bastante mayor que los tres cuartos de ítems de selección múltiple incluidos en la prueba entera. De las preguntas abiertas de la subescala, una corresponde a un formato de respuesta corta (Pregunta 6), y la otra a un formato de respuesta extensa (Pregunta 21).

La Pregunta 6 se ubica entre el tercio de preguntas más fáciles de la subescala. Este ítem se ajusta perfectamente bien al perfil de rendimiento nacional, sin mostrar ningún comportamiento diferencial en comparación a las demás preguntas de selección múltiple. Una situación muy distinta se observa en la Pregunta 21, la única de respuesta extensa. Este ítem se “cae” de la línea, presentando un porcentaje de respuestas correctas considerablemente menor que el resto.

Considerando que la Pregunta 21 es la única con un formato de respuesta extensa, la hipótesis de una interacción país-formato surge de inmediato. Si bien la Pregunta 6 también es una pregunta abierta, su estructura es considerablemente más simple que la de la Pregunta 21. De hecho, la Pregunta 6 sólo requiere escribir un número en un renglón. La Pregunta 21 es una de las preguntas publicadas por TIMSS, razón por la cual se reproduce a continuación. Como se puede observar en la figura 3, la complejidad de esta pregunta es considerable.

Figura 3

Ejemplo de Ítem de Respuesta Extensa – Pregunta 21

Juan Planea suscribirse a 24 números de una revista. El lee los siguientes avisos de dos revistas. *Ceds* son las unidades monetarias en el país de Juan.

<p>Revista Zona Joven</p> <p>24 números Primeros cuatro números GRATIS El resto 3 ceds cada uno.</p>	<p>Revista Lolomanía</p> <p>24 números Primeros seis números GRATIS El resto 3,5 ceds cada uno.</p>
---	--

¿Cuál revista es la más barata por los 24 números? ¿Cuánto más barata es?
Desarrolla tu respuesta.

El análisis directo de este ítem hace preguntarse además por su clasificación en el área de representación de datos. ¿Qué tarea es más demandante: entender el mensaje de los dos avisos publicitarios o realizar las operaciones necesarias para responder al enunciado? Si la respuesta es la segunda, esta pregunta podría estar mejor clasificada en fracciones y sentido numérico.

Como comentario final es interesante hacer notar que la Pregunta 21 fue no sólo para Chile, sino que para todos los países TIMSS, una de las preguntas más difíciles de la prueba. En el informe internacional se la usa para describir el tipo de problemas que típicamente fueron capaces de resolver el 10% de los alumnos con mejor rendimiento (Mullis *et al.* 2000: 61).

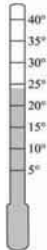
Mientras la Pregunta 21 da cuenta de lo que son capaces de hacer los alumnos con mayores habilidades matemáticas, la Pregunta 3 –la otra pregunta de la subescala publicada– da cuenta del tipo de problemas que los alumnos con menores habilidades fue capaz de resolver (Mullis *et al.* 2000: 85). Como se observa en la siguiente figura, la Pregunta 3 es un ítem de selección múltiple que requiere una lectura directa de los datos en una tabla.

Figura 4

Ejemplo de Item de Selección Múltiple – Pregunta 3

Esta tabla muestra la temperatura a distintas horas en cuatro días.

TEMPERATURA					
	6 A.M.	9 A.M.	Mediodía	3 P.M.	6 P.M.
Lunes	15°	17°	24°	21°	16°
Martes	20°	16°	15°	10°	9°
Miércoles	8°	14°	16°	19°	15°
Jueves	8°	11°	19°	26°	20°



Termómetro

¿Qué día y a qué hora la temperatura fue igual a la que se muestra en el termómetro?

A. Lunes, a Mediodía
 B. Martes, 6 A.M.
 C. Miércoles, 3 P.M.
 D. Jueves, 3 P.M.

M022260

Los alumnos chilenos obtuvieron 67,40% de respuestas correctas en esta pregunta, mientras que el promedio internacional fue de 79,20%. La diferencia entre Chile y el promedio internacional es algo menor que la diferencia promedio que separa a ambos grupos (11,80 versus 17,25 puntos porcentuales).

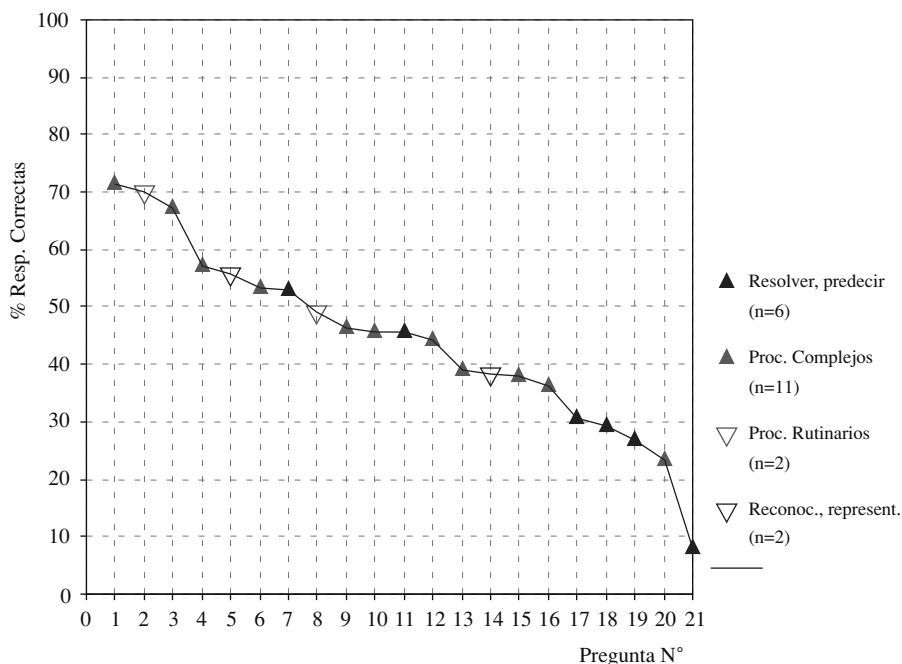
Hipótesis sobre las Expectativas de Rendimiento. ¿Presentan los ítemes que intentan medir las mismas habilidades cognitivas un nivel de dificultad similar? La figura 5 muestra el perfil de rendimiento de Chile en la subescala de representación de datos; distintos marcadores se utilizaron para dar cuenta de las distintas habilidades evaluadas por cada uno de ellos.

De acuerdo a la figura, las preguntas que demandan el uso de procedimientos rutinarios cubren un amplio espectro de dificultad, desde la Pregunta 1 a la 20. El otro grupo de preguntas que mide habilidades de orden superior (resolver y predecir) se encuentra algo más aglutinado hacia el sector de las preguntas de mayor dificultad (valor $p < 0,31$). Este grupo está constituido por las Preguntas 21, 19, 18 y 17. De igual modo, los ítemes que miden la categoría de recordar y representar, así como realizar procedimientos rutinarios, se encuentran algo más agrupados hacia los ítemes más fáciles, con valores $p > 0,39$.

Estos resultados se ajustan a lo esperado desde la perspectiva del desarrollo: las preguntas que miden habilidades cognitivas superiores presentan los valores p más bajos, y las preguntas que miden habilidades cognitivas básicas presentan los valores p más altos. Hablando desde una perspectiva educativa, es necesario dejar en claro que el antiguo currículum chileno de octavo básico (así como el currículum de la mayoría de

Figura 5

Perfil de Rendimiento según Expectativas de Rendimiento



Fuente: Base de datos TIMSS 1999.

los países en este nivel) ponía el énfasis en el manejo de destrezas básicas y la comprensión de problemas matemáticos (Ministerio de Educación 1999). Dado que la secuencia curricular refleja los pasos del desarrollo cognitivo, es imposible separar el efecto curricular del efecto propio del desarrollo.

Más allá de este énfasis en destrezas básicas, el antiguo currículum nacional cubría un amplio repertorio de habilidades. Conocer, representar, realizar procedimientos rutinarios, realizar procedimientos más complejos, comprender problemas matemáticos, resolver problemas y comunicarse, todos eran mencionados en el marco curricular. La única habilidad evaluada en la prueba TIMSS y no mencionada en las intenciones curriculares era predecir (Ministerio de Educación 1999). ¿Son las preguntas que exigen predecir de mayor complejidad que las preguntas que evalúan otras destrezas cognitivas? La subescala aquí analizada sólo incluye dos preguntas del tipo predecir, por lo que resulta difícil generalizar. Sin embargo, ninguna de estas preguntas se encuentra entre las más complejas.

Los libros de texto pueden entenderse como el vínculo entre currículum intencionado y currículum implementado. Sirven de interfaz entre los objetivos curriculares y las lecciones que se llevan a cabo en la sala de clases. En Chile, los textos escolares son ampliamente usados por los profesores, siendo para muchos el currículum *de facto* al

momento de preparar clases o enseñar. Asumiendo que los contenidos y destrezas incluidos en los textos escolares tienen mayores probabilidades de ser enseñados a los alumnos, es de interés conocer hasta qué punto las expectativas de rendimiento evaluadas en TIMSS formaban parte de los textos de matemática utilizados por los alumnos de octavo básico el año de la prueba.

Siguiendo las especificaciones curriculares, los textos chilenos asignaban una cantidad importante de su espacio al desarrollo de habilidades básicas como conocer (41-50%) y usar procedimientos rutinarios (31-40%)⁸. Habilidades de orden superior tales como investigar, o resolución de problemas y razonamiento matemático, solo aparecían en 11-20% y 1-10% del espacio disponible (Ministerio de Educación 1999). Estos énfasis apoyan la adquisición de destrezas cognitivas básicas, y son consistentes con los resultados presentados en la figura anterior.

Otro foco de análisis se refiere a la cantidad de preguntas que miden cada habilidad en la subescala. Tal como lo indica la figura 5, la distribución de ítems por categoría cognitiva es dispareja. Mientras en las dos categorías de más alta demanda cognitiva hay seis ítems del tipo resolver y predecir y once ítems que requieren usar procedimientos complejos, en las categorías inferiores hay sólo dos preguntas de recordar y representar y otras dos en usar procedimientos rutinarios.

Esta desigual distribución de preguntas según categoría cognitiva puede justificarse por el mayor énfasis de TIMSS en la evaluación de habilidades cognitivas superiores. No obstante, es necesario hacer notar que la distribución de preguntas según categoría cognitiva es mucho más homogénea si se mira la prueba de matemática entera, en comparación con la subescala aquí analizada.

Hipótesis sobre Subáreas de Contenido. El reporte internacional menciona tres subáreas de contenido que conforman la subescala de representación de datos (Mullis *et al.* 2000: 93):

- Representación e interpretación de datos en tablas y gráficos.
- Rangos y promedios.
- Probabilidades simples (cálculo y comprensión).

Con el objeto de probar si las preguntas que evalúan la mismas subáreas de contenido presentan un nivel de dificultad similar, las 21 preguntas que conforman la subescala de representación de datos fueron graficadas utilizando distintos marcadores según el contenido evaluado por cada una de ellas (figura 6).

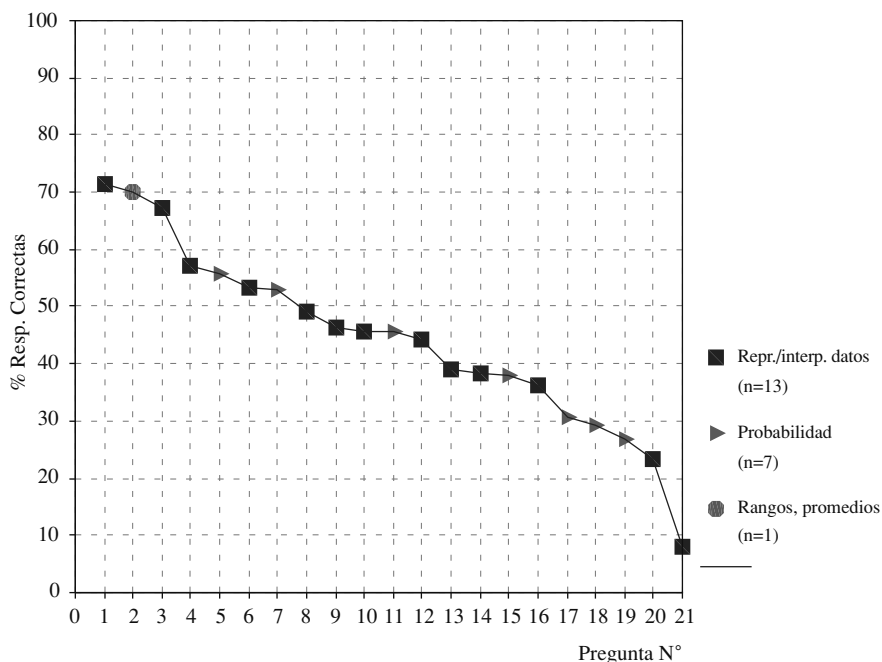
De acuerdo a la figura, la cantidad de preguntas varía considerablemente dependiendo del tópico. Mientras en representación e interpretación de datos hay 13 ítems, en rangos y promedios sólo hay uno. El tópico de probabilidades ocupa una posición intermedia, con siete preguntas a su haber.

La Pregunta 2 es la única que evalúa la categoría rangos y promedios, y resultó ser una de las fáciles de la subescala, con 70% de respuestas correctas. Los ítems de

⁸ La suma de los bloques de un texto escolar es mayor que 100% dado que un mismo bloque podía ser codificado como estimulando una o más habilidades. Detalles sobre la metodología utilizada pueden encontrarse en Schmidt, Raizen, Britton, Bianchi & Wolfe (1997).

Figura 6

Perfil de Rendimiento según Subáreas de Contenido



Fuente: Base de datos TIMSS 1999.

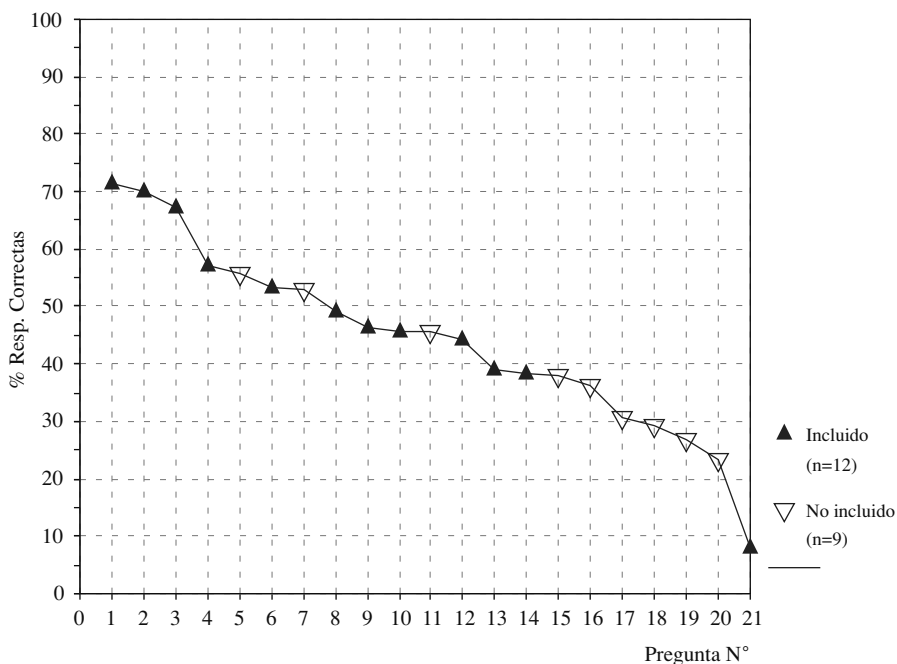
representación e interpretación de datos se encuentran distribuidos en todo el continuo de rendimiento; de hecho, tanto la pregunta más fácil como la más difícil corresponden a esta categoría. Las preguntas de probabilidades se encuentran algo más concentradas hacia el extremo de mayor dificultad de la línea. Esto es especialmente evidente para las Preguntas 17, 18 y 19. Como se verá en la siguiente sección, este patrón puede ser fácilmente vinculado a los contenidos incluidos en el antiguo currículum de matemáticas.

Hipótesis sobre Intenciones Curriculares. Cada pregunta de la subescala de representación de datos fue clasificada de acuerdo a las especificaciones del antiguo currículum nacional. Si, de acuerdo al currículum, se esperaba que los alumnos de octavo básico manejaran los contenidos y habilidades necesarios para responder correctamente a una pregunta, la pregunta se marcó como “incluida” en el currículum. Si, por el contrario, la pregunta iba más allá de lo esperable para un alumno de octavo de acuerdo al currículum, la pregunta se marcó como “no incluida”.

La idea entonces era ver si las preguntas incluidas en el currículum se agrupaban entre las más fáciles, y las no incluidas entre las más difíciles. Con este objetivo, cada una de las preguntas de la subescala fue marcada de acuerdo a su estatus curricular (ver figura 7).

Figura 7

Perfil de Rendimiento según Intenciones Curriculares



Fuente: Base de datos TIMSS 1999.

Un claro patrón de desempeño emerge esta vez. De acuerdo a lo esperado, las preguntas incluidas en el currículum se agrupan hacia el lado de las preguntas más fáciles, mientras que las preguntas no incluidas tienden a aglutinarse hacia la zona de mayor dificultad. Así es como las preguntas incluidas en el currículum dominan desde la Pregunta 1 a la 14; sólo tres ítemes no intencionados se ubican dentro de este rango (Preguntas 5, 7 y 11).

Contra lo esperado, el ítem más difícil de todos es una pregunta incluida en el currículum (Pregunta 21). Una vez más, ésta es la pregunta de respuesta extensa cuya clasificación en representación de datos fue cuestionada. Tomando en cuenta ambos factores (formato y tópico) es posible explicarse el porqué esta pregunta se encuentra tan separada del resto de los otros ítemes incluidos en el currículum nacional.

Centrándose ahora en las preguntas no incluidas en el currículum, éstas se ubican mayoritariamente entre las Preguntas 15 y 20, y presentan un porcentaje de respuestas correctas bajo el 39%. Se distancian de este grupo las Preguntas 5, 7 y 11, que presentan puntajes por sobre el promedio nacional de 44% de respuestas correctas.

Si se comparan las figuras 6 y 7, se hace evidente que los ítemes no incluidos en el currículum son todas las preguntas de probabilidades, y dos preguntas de representación e interpretación de datos. Probabilidades es, de hecho, un tópico que el antiguo currícu-

lum sólo introducía en cuarto medio. No obstante, representación e interpretación de datos es un tópico que el currículum introduce desde 6^{to} a 8^{vo} básico (Ministerio de Educación 1999). ¿Por qué dos preguntas (Preguntas 16 y 20) de esta subárea de contenido fueron clasificadas como no incluidas en el currículum siendo que ninguna de ellas requería predecir (la única habilidad no incluida en el currículum)? Si bien es posible que estas preguntas incluyeran algún tipo de gráfico o tópico que fuera más allá de las expectativas curriculares para 8^{vo} básico, este tipo de inconsistencias levanta dudas sobre la confiabilidad de los datos⁹.

De acuerdo a las intenciones curriculares, los demás 11 ítemes de representación e interpretación de datos aparecen como “incluidos” en la figura 7, así como también la única pregunta que evalúa rangos y promedios (Pregunta 2). Rangos y promedios es un tópico que se introduce en el primer ciclo de enseñanza básica y se refuerza hasta 8^{vo} básico (Ministerio de Educación 1999). Estas intenciones curriculares respaldan el alto porcentaje de respuestas correctas (70%) obtenidas en esta pregunta.

Con el objeto de verificar si existía una diferencia significativa en el promedio de respuestas correctas para ambos grupos de preguntas (incluidas y no incluidas en el currículum), se realizó una prueba *t* para muestras independientes. La Pregunta 21 fue excluida del análisis ya que había suficientes argumentos (formato y tópico de la pregunta) para explicar su peculiar comportamiento. Calculado de este modo, las preguntas incluidas en el currículum tenían un promedio de respuestas correctas de 53%, comparado con un 37,64% obtenido por las preguntas no incluidas ($p < 0,01$). Como se verá más adelante, estos resultados no siguen la tendencia que reporta el informe internacional cuando se refiere al conjunto de los países.

Hipótesis sobre el Currículum Implementado. En la figura 8 se presentan dos porcentajes de respuestas correctas por pregunta: uno corresponde al porcentaje obtenido por los alumnos que ya pasaron la materia, y el otro al porcentaje obtenido por los alumnos que aún no han pasado la materia. Ambos grupos se conformaron de acuerdo al reporte de los profesores de los alumnos evaluados. La comparación de ambos porcentajes permite ver si, de acuerdo a lo esperado, los alumnos que ya pasaron la materia obtienen mejores resultados que los alumnos que aún no la pasan.

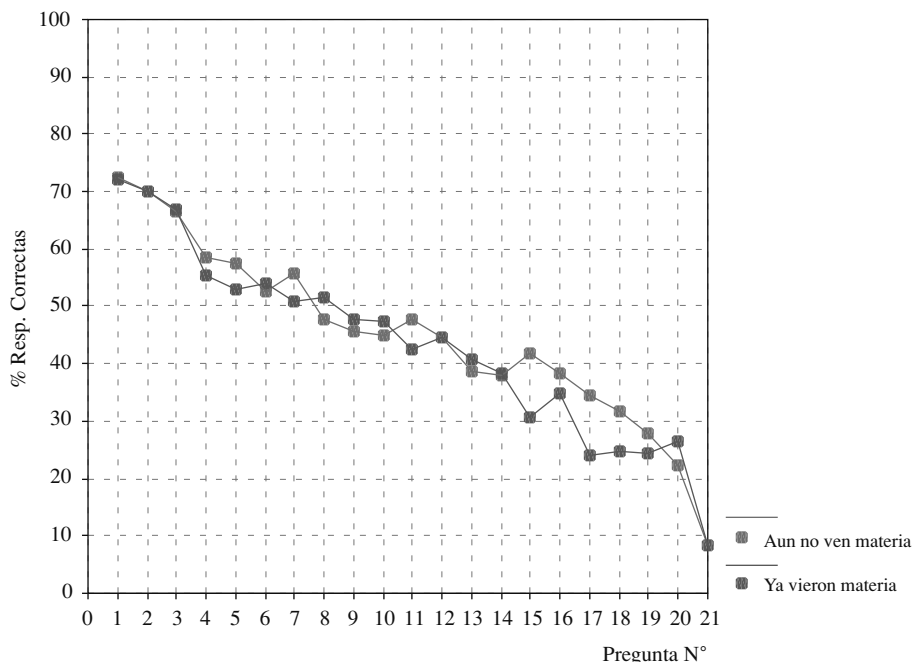
La figura muestra que no existen diferencias importantes entre los alumnos que ya pasaron la materia y los que no. Para ambos grupos, el perfil de rendimiento es bastante similar, con pequeñas variaciones de un ítem a otro. En el mejor de los casos, estas diferencias favorecen a los alumnos que ya pasaron la materia. De hecho, este fue el caso en 10 de los 21 ítemes de la subescala. En las restantes 11 preguntas, los alumnos que aún no pasaban la materia obtuvieron mejores resultados que sus pares más avanzados en el programa de estudio. Si bien estas diferencias no son significativas, los alumnos que aún no pasaban la materia obtuvieron en promedio mejores resultados que los que ya la habían pasado (45,01% versus 43,26%, respectivamente).

Llama la atención que en las siete preguntas de probabilidades (Preguntas 5, 7, 11, 15, 17, 18 y 19) les fue mejor a los alumnos que aún no pasaban la materia. Las mayores

⁹ Las preguntas de la prueba fueron marcadas como “incluidas” o “no incluidas” en el currículum por un especialista en matemáticas, quien trabajó teniendo como referencia el antiguo marco curricular.

Figura 8

Perfil de Rendimiento según Grado de Implementación Curricular



Fuente: Base de datos TIMSS 1999.

diferencias de rendimiento también se observan en estas preguntas, en donde el grupo aún no enseñado superó por entre 11% y 3,70% al otro grupo.

CONCLUSIONES Y DISCUSION

Este estudio profundiza en el análisis de una de las subescalas de la prueba TIMSS: representación de datos. Su objetivo es comprender las razones que subyacen al patrón de desempeño observado en el caso de Chile.

Para los alumnos chilenos –y no así para los del resto del mundo– el ítem que presentó mayor dificultad fue, de lejos, la Pregunta 21, una pregunta con formato de respuesta extensa. Un antecedente importante de considerar es que TIMSS 1999 fue la primera experiencia de prueba estandarizada de matemática, con preguntas abiertas, que tuvieron estos alumnos¹⁰. O’Leary (2001) comparó los resultados relativos obtenidos por distintos países según el formato de las preguntas. Los alumnos irlandeses obtienen

¹⁰ La prueba SIMCE del año 2000 fue la primera en incluir preguntas abiertas en la prueba de matemática de 8^{vo} básico.

mejores resultados en las preguntas de respuesta extensa, lo que es consistente con la tradición de pruebas de ensayo utilizadas en ese país. El autor concluye que la frecuencia de exposición de los alumnos a distintos formatos de pregunta explica en forma importante el desempeño.

Resumiendo, la evidencia sugiere un *efecto de interacción formato-país*, en donde los ítems presentan un Funcionamiento Diferencial (DIF) dependiendo del contexto educativo y formato de las preguntas. Esto quiere decir que la probabilidad de responder correctamente a una pregunta no estaría únicamente condicionada por las habilidades matemáticas de los alumnos, sino además por su familiaridad con los distintos formatos de preguntas. Análisis psicométricos tipo DIF arrojarían más luces sobre este aspecto de la evaluación.

Si este efecto de interacción es real, entonces es inevitable preguntarse sobre la validez de las comparaciones entre distintas áreas de contenido. ¿Es por casualidad que álgebra, la subescala con la mayor proporción de ítems de respuesta extensa (7 de 35), es también la subescala con los peores resultados para Chile? Para evitar este tipo de incertidumbres es recomendable mantener relativamente estable la proporción de ítems con distinto formato utilizados en cada una de las subescalas.

El análisis por *destrezas cognitivas* (expectativas de rendimiento) muestra que las preguntas que evalúan habilidades cognitivas superiores son algo más difíciles de responder que las que se centran en la evaluación de destrezas básicas. Estos resultados son consistentes con las etapas del desarrollo cognitivo y también con los énfasis del antiguo marco curricular.

Centrándose ahora en las *subáreas de contenido* incluidas en representación de datos (rangos y promedios, representación e interpretación de datos, probabilidades), se observa una desigual distribución de las preguntas según estos tres tópicos. Claramente, contar con sólo una pregunta de rangos y promedios versus 13 de representación e interpretación de datos, no parece ser la mejor solución. Una distribución más equilibrada de las preguntas según tópicos mejoraría la calidad de la prueba y potenciaría las interpretaciones ligadas a aspectos curriculares.

El análisis de las *preguntas publicadas* levanta dudas sobre su clasificación en las distintas áreas de contenido. Se podría argumentar que para que la Pregunta 21 correspondiera a una pregunta de representación e interpretación de datos, ésta debería indagar sobre la información explícitamente presentada en los avisos y no en las inferencias a las que los alumnos pueden llegar por medio de operaciones matemáticas.

Considerando que Chile obtuvo uno de sus más bajos resultados en la categoría fracciones y sentido numérico, uno está inclinado a pensar que la presencia de operaciones matemáticas en la Pregunta 21 está dando cuenta, al menos en parte, de la dificultad presentada por este ítem.

El análisis de las preguntas según *intenciones curriculares* fue de gran utilidad para comprender mejor el perfil de desempeño en la subescala de representación de datos. De acuerdo a lo esperado, los alumnos chilenos obtuvieron mejores resultados en aquellas preguntas “incluidas” en el currículum (después de excluir el único ítem de respuesta extensa). Sin embargo, estos resultados no son consistentes con el reporte internacional, en donde se concluye que no hay un efecto curricular cuando se analiza la prueba como un todo.

Cabe señalar que en este documento Chile aparece con un promedio de 31% de respuestas correctas para todas las preguntas de la prueba de matemática, mientras que el promedio de respuestas correctas para las preguntas incluidas en el currículum sólo llega a 32% (Mullis *et al.* p. 350). Una situación similar se reporta en el caso de los demás países. Los rankings basados en todas las preguntas de la prueba o sólo en las preguntas intencionadas en los currículos de los países no muestran mayores diferencias en la posición relativa de los mismos¹¹. “Es evidente que la selección de las preguntas no tiene un mayor impacto en la posición relativa de los países” (Mullis *et al.* 2000: 349).

Estas discrepancias entre los resultados de Chile en una subescala, por una parte, y los resultados en la prueba entera (tanto para Chile como para los demás países) pueden explicarse de la siguiente manera. En la subescala de representación de datos las preguntas incluidas en el currículum son básicamente las preguntas de representación e interpretación de datos, mientras que las preguntas no intencionadas son las preguntas de probabilidades. Mi hipótesis es que ambos grupos de preguntas en realidad evalúan el manejo de tópicos que se aprenden de formas muy distintas.

Representación e interpretación de datos es un tema de “sentido común”; y los alumnos aprenden sus rudimentos ya sea en otras asignaturas, o en los medios de comunicación masiva, fundamentalmente diario y televisión (Howson 2001). El tópico de probabilidades, por el contrario, es algo que se aprende principalmente en el colegio y en el ramo de matemáticas. Mi hipótesis es que el mejor desempeño de los alumnos chilenos en la subescala de representación e interpretación de datos se debe más a sus experiencias “extraescolares” con el tema, y no tanto a la instrucción recibida en la escuela.

La Pregunta 5 es interesante en este sentido. Si bien ésta es una pregunta de probabilidades, el problema que presenta es extremadamente familiar para los alumnos: la probabilidad de obtener cara o sello al lanzar una moneda al aire. Los alumnos chilenos respondieron inesperadamente bien en esta pregunta, igualando el promedio internacional. Estos resultados apoyan la idea de un fenómeno de aprendizaje informal que explicaría estos buenos resultados.

Esta hipótesis también es consistente con los resultados del *análisis de implementación curricular*. Aquí se descubre que no hay diferencias de rendimiento entre aquellos alumnos que ya habían pasado los contenidos de la subescala de representación de datos y aquellos que no. Incluso, los alumnos que ya habían pasado la materia obtuvieron resultados ligeramente peores que aquellos que aún no la pasaban. Estos resultados son desilusionantes, y llevan a preguntarse por la efectividad de la educación chilena, al menos en el área de representación de datos. También llevan a preguntarse por la magnitud del efecto instruccional en otras áreas de contenido y en otros países.

A modo de cierre, un breve comentario sobre las *discrepancias observadas entre currículum intencionado y currículum implementado*. De acuerdo al antiguo currículum chileno, probabilidades era un tema que debía ser introducido en IV medio. Sin embargo, el reporte de los profesores indica que el 35% de los alumnos de 8^{vo} básico ya había

¹¹ Llama la atención que Chile fuera el país con la menor cantidad de preguntas (puntos) marcadas como “apropiadas” de acuerdo al marco curricular. Sólo 58% (98 puntos) de los 162 ítemes (169 puntos) de la prueba fueron identificados como tales en el procedimiento de Análisis Curricular que se llevó a cabo.

visto esta materia. Por otra parte, tenemos que representación e interpretación de datos era una materia que se pasaba desde 6^{to} a 8^{vo} básico, pero sólo el 49% de los alumnos de 8^{vo} la había visto. El mismo porcentaje de alumnos ya había visto promedio aritmético, una materia que debería pasarse bastante antes de llegar a 8^{vo} básico (Mullis *et al.* p. 179).

¿Qué tipo de guía pedagógica estaba ofreciendo el antiguo currículum de matemática? ¿Por qué los profesores no seguían sus lineamientos? Una posible explicación es lo anticuado del programa de estudio. La enseñanza de matemáticas ha cambiando mucho desde 1980, el año de estreno de este documento. Es razonable pensar que los profesores no lo seguían para dar cabida a nuevas tendencias curriculares. De hecho, estas nuevas tendencias fueron bastante promovidas y discutidas durante el desarrollo del nuevo currículum. Si bien los nuevos marcos de estudio estaban listos desde 1996, su implementación en la sala de clases se demoró hasta el 2002, año en que estuvieron listos los nuevos programas de estudio. De ahí que sea razonable pensar que durante estos años imperó bastante confusión con respecto a qué currículum utilizar: el antiguo o el nuevo. Ahora que el nuevo está en plena vigencia, con énfasis se recomienda monitorear su implementación en las salas de clases. De este modo se podrán apoyar mejor los procesos de enseñanza-aprendizaje en la sala de clases.

REFERENCIAS

- GONZALEZ, E. J., J. A. MILES. (Eds.). (2001). *TIMSS 1999 User Guide for the International Database* [Manual de Uso Base de Datos Internacional TIMSS 1999]. Chestnut Hill, MA: International Study Center Lynch School of Education Boston College.
- HOWSON, G. (2001). *TIMSS, common sense and the curriculum* [TIMSS, el Sentido Común y el Currículum]. Manuscrito enviado a publicación.
- LARA, M., M. LUQUE, A. MENDOZA. (1998). *Matemática 8*. Santiago de Chile: Editorial Universitaria.
- MARTIN, M. O., K. D. GREGORY, S. E. STEMLER. (2000). *TIMSS 1999 Technical Report: IEA's Repeat of the Third International Mathematics and Science Study at the Eight Grade* [Informe Técnico TIMSS 1999: Repetición del Tercer Estudio Internacional de Matemáticas y Ciencias para 8^{vo} básico de IEA]. Chestnut Hill, MA: Boston College.
- MINISTERIO DE EDUCACION. (1980). *Planes y Programas de la Educación General Básica*. Santiago de Chile: Ministerio de Educación.
- MINISTERIO DE EDUCACION. (1999). [TIMSS 1999 análisis curricular]. Archivo de datos sin publicar.
- MULLIS, I. V. S., M. O. MARTIN, E. J. GONZALEZ, K. D. GREGORY, R. A. GARDEN, K. M. O'CONNOR, S. J. CHROSTOWSKI, T. A. SMITH. (2000). *TIMSS 1999 International Mathematics Report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eight Grade* [Reporte Internacional TIMSS 1999 Matemáticas: Repetición del Tercer Estudio Internacional de Matemáticas y Ciencias para 8^{vo} básico de IEA]. Chestnut Hill, MA: Boston College.
- O'LEARY, M. (2001). *Item format as a factor affecting the relative standing of countries in the Third International Mathematics and Science Study (TIMSS)* [Formato de los ítemes como un factor explicativo del rendimiento relativo de los países en el Tercer Estudio Internacional de Matemáticas y Ciencias (TIMSS)]. Artículo presentado en 2001 la Conferencia anual de la American Educational Research Association, Seattle, WA.

- TIMSS 1999 International Database 1999 Assessment Data* [CD-ROM, Base de Datos Internacional TIMSS 1999]. (2001). Chestnut Hill, MA: International Study Center Lynch School of Education Boston College [Produce y Distribuye].
- SCHMIDT, W. H., S. A. RAIZEN, E. D. BRITTON, L. J. R. G. WOLFE. (1997). *Many Visions, Many Aims Volume 2: A cross-national investigation of curricular intentions in school mathematics* [Visiones y Aspiraciones Volumen 2: un estudio internacional de intenciones curriculares en matemáticas]. Dordrecht, The Netherlands: Kluwer Academic Publisher.