INVESTIGACIONES

PROFILE OF CHILEAN ACHIEVEMENT IN THE TIMSS 1999 DATA REPRESENTATION SUB-SCALE

Perfil de rendimiento de Chile en la Subescala de Representación
de Datos TIMSS 1999

Prof. María José Ramírez*

International Study Center, Boston College, Lynch School of Education, Beacon 188,
Chestnut Hill, MA 02467, U.S.A. E-mail: ramiremb@bc.edu

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 education.

Resumen

Por primera vez, Chile participó en el Estudio de Tendencias en Matemáticas y Ciencias de IEA (TIMSS 1999), evaluando los conocimientos de una muestra representativa a nivel nacional de 5907 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.

INTRODUCTION

In TIMSS 1999, Chile ranked 35 out of 38 countries in the mathematics test. Its average achievement was 392 scale points[1], and its performance comparable to that of countries such as Turkey, Jordan, Iran, and Indonesia (Mullis *et al.* 2000, chap. 1). The mathematics test included five content areas: "fraction & number sense", "measurement", "data representation", "geometry", and "algebra". In all five areas the Chilean performance was among the poorest of the participating countries. However, significant differences were reported in the relative achievement in these areas. While the weakest results for Chile were obtained in algebra and fraction & number sense, the strongest results were obtained in data representation (Mullis *et al.* 2000, chap. 3).

What could explain this better performance in data representation in comparison to the other content areas? Is there something to be learned about this topic that could help improve the performance in the other topics? Answering these questions would strictly imply running a comparative analysis among the different content areas. However, a good starting point would be to look deeply into one area: the data representation sub-scale.

The analyses were guided by five major concerns/hypotheses raised in the literature as affecting students' learning and their performance in standardized tests:

1. *Item formats hypothesis.* Do the different item formats presented in the test help to explain the pattern of achievement in the data representation questions? How does student achievement vary according to this format? Are the open-ended questions harder than the multiple-choice items?
2. *Performance expectations hypothesis*. Can the gradient of achievement be explained by the underlying performance expectations (cognitive skills in TIMSS jargon) required to answer each item? Do the most difficult items require the use of higher order skills to be answered? Do easier items correspond to lower order skills items?
3. *Content sub-areas hypothesis*. Is it possible to explain the Chilean achievement profile according to different content sub-areas assessed within the data representation sub-scale? Do the items of the same content sub-area cluster together in the achievement profile line?
4. *Intended curriculum hypothesis*. Intended curriculum refers to the curricular goals and intentions that specify what the students are expected to learn at school. The official curriculum is the prototype of intended curriculum document. Thus, the research question turns to: Were the items expected to be taught to 8th grade students easier than those that were not?
5. *Implemented curriculum hypothesis.* The implemented curriculum refers to the instruction that is actually provided by the teachers in the school setting. The question that raised is: Do the students who were taught the topics required to answer the items show a better performance compared to those who were not yet taught the topics?

The following sections explain how each of these questions were addressed, and present the major findings of the study. Those results are then discussed in a broader perspective in the final chapter of this paper.

---

[1]   The TIMSS scale has a mean 500 and a standard deviation of 100.

METHOD

Sources of information. In Chile, a sample of 5907 eighth graders from 185 schools took the TIMSS test in November 1998[2]. Their math teachers completed a questionnaire, providing information about their background and preparation to teach, topics covered, and instructional practices implemented, among others. The Chilean national research coordination[3] provided information about the educational system as a whole, curricular framework and emphases, nature of country-level assessments, among other contextual information.

All this information was captured in the *TIMSS 1999 International Database* (2001), which constitutes the main source of information for the study presented here. Occasionally data was directly retrieved from the *TIMSS 1999 International Mathematics Report* (Mullis *et al.* 2000); a reference is provided in these cases. Details about how this information was collected and the technical procedures used in its production can be found in the TIMSS 1999 publication series (Mullis *et al.* 2000; Martin, Gregory & Stemler 2000; González & Miles 2001).

Data from the *TIMSS Curricular Analysis Study* was also used in this report. Even though Chile did not participate in the original study (Schmidt, Raizen, Britton, Bianchi & Wolfe 1997), the same methodology was applied in a data collection effort carried on in 1999. Basically, the method consisted of an analytical approach where curricular documents and textbooks were coded according to the content and performance expectations presented in each "block". Blocks were the basic unit of analysis in the documents, usually ranging in length from a simple statement (e.g. a curricular standard) to a short paragraph (e.g. an instructional activity). This information was used to support statements about the curricular intentions and the most likely math experiences to which 8th grade students have been exposed. In Chile, the mathematics textbook used by 90% of eight graders was coded, as well as the Chilean national curriculum in use the year TIMSS assessment took place[4] (Ministerio de Educación 1999). Since this curricular framework was replaced by a new one in the school year that started in March 2002, we will refer to it as the *old* curriculum/framework.

Procedure. To better appreciate the Chilean performance in the 21 data representation items under analysis, these items were numbered from 1 to 21, according to their difficulty level for the Chilean students. The difficulty level was based on the mean percent of correct answers obtained in each item (*p* value). The easiest item –the one correctly answered by the highest percentage of Chilean students– was labeled Item 1; the second item with the highest percentage of correct answers was labeled Item 2, and so on until the hardest item was labeled Item 21. Across all the figures shown in this report, the item numbers always refer to the same question.

---

[2] Chile as well as the other south-hemisphere countries took the test at the end of the school year of 1998; north-hemisphere countries collected their data during 1999.

[3] The Chilean national research coordination for TIMSS 1999 was under the direction of the Ministry of Education, Programa SIMCE, unidad de Estudios Internacionales.

[4] The Chilean curricular framework used in the curricular analysis study was *Planes y Programas de la Educación General Básica* (Ministerio de Educación, 1980). The textbook sampled was *Matemática 8* (Lara, Luque & Mendoza, 1998).

The items were then sorted from easiest to hardest, and graphed in a scatterplot with Items 1 to 21 presented in the x-axis and the percent of correct responses in the y-axis. A data representation achievement profile was then obtained, with the item plots forming a smooth slope of difficulty level for the Chilean performance. This same procedure was repeated several times in different figures, just changing the plots symbols in order to represent a third variable. This third variable was directly related to each one of the hypotheses under analysis: item format, performance expectations, content sub-areas, curricular intentions and curricular implementation.

The Chilean and international percents of correct responses for the 21 items under analysis were directly taken from the TIMSS data almanac files. Information about the multiple categorization systems used to describe the items of the test – format, cognitive skills, content sub-areas, and curricular intentions – was also directly retrieved from the item information files contained in the TIMSS database.

The classification of items by cognitive skill required some reorganization of the original categories. The international report mentions five performance expectations: "knowing", "using routine procedures", "investigating & problem solving", "mathematical reasoning", and "communicating" (Mullis *et al.* 2000: 318). The TIMSS database provides a more specific classification system, consisting of 10 categories: "recalling mathematical objects & properties", "representing", "performing routine procedures", "using more complex procedures", "solving", "predicting", "solving, describing & discussing", "formulating & clarifying problems and situations", "generalizing", and "recognizing equivalents". All but the last three categories apply to data representation items.

For the purposes of this report, the second classification system was used because of its direct relation with the tasks presented in the data representation items. However, some changes were introduced in order to make the classification more meaningful for the content area under analysis. Specifically, five categories were rearranged into only two categories: "recalling mathematical objects & properties" was merged with "representing" into one group; "solving" was merged with "predicting" and with "solving, describing & discussing", into another group.

The computations of the percent of correct responses for students who were taught the topic and those who were not yet taught the topic –curriculum implementation status– required the combination of students' achievement scores and data from the teachers' questionnaire. Teachers were asked when students in their mathematics class were taught each topic in the tests. In the "Data representation, analysis and probability" section, they were specifically asked about the three sub-content areas that form the data representation sub-scale:

– Representation & interpretation of data in graphs, charts, and tables.
– Arithmetic mean[5].
– Simple probabilities – understanding and calculations.

---

[5]   Teachers were asked about arithmetic mean instead of means & ranges – the official label of this sub-content area. Most probably this change was introduced in order to reflect in a more accurate way what is actually measured by Item 2 (the only item belonging to this sub-content area) that is to compare two average scores.

For each of these topics, teachers were asked if the topic was: a) Taught before this year; b) Taught 1 to 5 periods this year; c) Taught more than 5 periods this year; d) Not yet taught; e) Do not know. Their answers were reclassified into two dichotomous groups: already taught (which includes points one to three) and not yet taught (which included the fourth point only). The "Do not know" answers –that accounted for 5% of the responses– were excluded from the analyses.

In each sub-content area, students were classified according to their teachers' answers. For example, a student could be reported as having been taught representation and interpretation of data and arithmetic mean, but not yet taught simple probabilities. In this case, this student was counted as "already taught" in the first two sub-areas, but as "not yet taught" probabilities. Then, within each sub-content area, separate percents of correct responses were calculated for every item.

RESULTS

Chilean Achievement Profile versus International Mean Achievement. Figure 1 presents the Chilean difficulty level (percent correct) of the 21 items that compose the data representation sub-scale. As previously explained, the items were sorted from the easiest (Item 1) to the hardest (Item 21), forming a uniform gradient for this country. The graph also shows the international mean difficulty level[6] for each item. The international mean was included in order to have an external criterion against which to compare the Chilean performance.

Several comments can be made about the graph above. First of all, it is worth noting the wide rage of Chilean difficulty levels for the data representation items. While 71.60% of the students answered Item 1 correctly, only 8.15% got the correct answer in Item 21[7]. If we compare the Chilean range of achievement against the international achievement, the former was more than one an a half times bigger than the latter – 63.45 versus 38.20 percent points of differences between the easiest and the hardest items for Chile and the international composite, respectively.

Secondly –and directly related to the previous paragraph– comparing the differences in item difficulty for Chile and the international segment, these differences tend to increase from the easiest to the hardest items. For example, while in Item 1 both lines are relatively close one to another (reflecting a difference of 6.60 points between Chile and the international average), in Item 21 the lines are much more apart (reflecting differences of 40.10 points). One clear exception to this pattern was Item 5, which presented almost the same difficulty level for Chile and the international composite – with 1.70 percent points of difference favoring the latter.

---

[6]   The international mean difficulty level was calculated from the percent of correct responses across the 38 countries participating in the study.

[7]   In the international report (Mullis *et al.* 2000: 65), the Chilean overall percent correct for Item 21 is 5%, and not 8,15% as here reported. The difference arises because the international reports only counted those students who get maximum score (2 points), while in this report were also counted those students who obtained partial credit (1 point). In the calculations, the latter were given half of the weight of the former to take into account the differences in score points.

Figure 1

Achievement Profile for Chile and International Composite
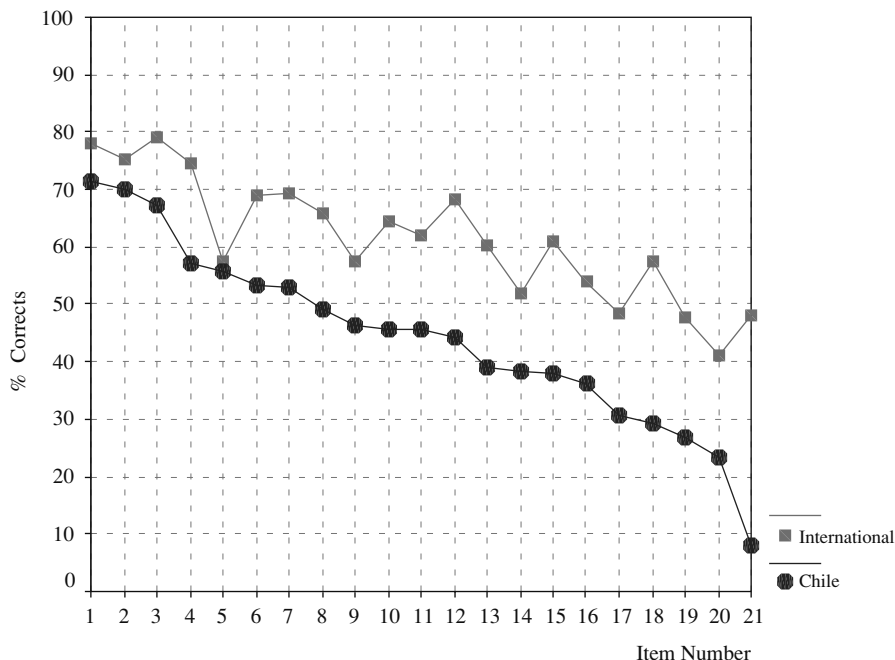
Source: TIMSS 1999 data files.

Is there something special about Item 5 that can explain why it is out of the pattern when comparing the Chilean and international profiles? Item 5 is a multiple-choice question that asks about the likely result of a fifth coin toss; it is a probability question (sub-content area) that 55.70% of the Chilean students answered correctly. Even though probability was not a topic intended to be taught to 8th grade students according to the old curriculum, it was one of the easiest items of the sub-scale.

Item 5 required the students to recall mathematical objects and properties, forming part of the more basic performance expectation level included in the test. It could be hypothesized that the Chilean performance equaled the international segment because of the basic cognitive skills demanded by this problem. But this is not a good explanation, because the item would be easier not only for the Chilean students but for all the students of the participating countries. A more plausible explanation refers to the familiarity of the problem presented in the item. That the probability of tossing a coin is always the same –no matter what the previous results were– is something very probable to be learned informally, as part of the "street mathematics" curriculum. This line of argumentation will be expanded in this paper.

Third, overall, most difficult items for Chilean students were also the most difficult items for the students across all TIMSS countries, and vice versa. What changed drasti-

cally between one group and the other was their mean achievement in data representation. This is graphically depicted by the distance separating both profile lines. On average, 17.25 percent points separate both lines. These are strong and significant differences ($p < .005$). This distance can also be depicted using the TIMSS scale score: while Chile obtained 429 points in the data representation sub-scale, the international mean was 487 points (Mullis *et al.* 2000: 97).

The fourth and final comment is devoted to Item 21. This question is definitely out of the slope pattern depicted by the Chilean line. On average, when passing from one item to another, the difficulty level changes two or three percent points. Nonetheless, in the case of Item 21, its percent of correct answers was 15 percent points less than the item with the next difficulty level. The international profile does not replicate this pattern. A somewhat drastic p value change is also observed when passing from Item 3 to Item 4 (with 10 percent points of difference); however, here the difference seems to respond to an international pattern followed by all the countries. As we will see in the next section, a plausible explanation for this observation is the special format of Item 21.

Testing the Item Format Hypothesis. The TIMSS 1999 test included three basic item formats (Martin, Gregory & Stemler 2000, chap. 3):

– *Multiple-choice items*. Students were asked to select their answer among four to five options presented below the stem. These items were worth one point, and constituted three-fourths of the questions of the math test (125 out of 162).
– *Short-answer items*. Students were expected to generate their answer, writing an operation, a number, and/or some kind of brief response. These items constituted a little more than 10% of the items of the test (21 out of 162), and they were worth one point.
– *Extended response items*. Students were expected to generate a more long and elaborate response, supporting their answer, writing whole sentences as their final response, or answering multi-step items with two or three sections. These items constituted a little less than 10% of the items of the tests (16 out of 162), and they were worth two or three points.

In order to see if there is a *format effect* that could explain the profile of Chilean achievement, the 21 data representation items were plotted using different marks according to their type of format (figure 2).

The figure shows that 19 out of 21 items in the data representation content area have a multiple-choice format. This is a substantially higher proportion compared to the three-fourths of items with this format in the entire test. Of the open-ended questions presented in this content area, only one corresponds to a short-answer type (Item 6), and another one to an extended response item (Item 21).

The only short-answer item belongs to the first third of item difficulty (group of easiest items). Item 6 fits perfectly well the profile line, showing no special pattern in relation to the rest of the multiple-choice items. A very different situation is observed in the case of the only extended-response question included in the sub-scale. Item 21 clearly falls below the projected line of the slope flow, presenting a percent of correct responses considerably lower than the rest of the items (8.15%). This drastic drop in the achievement profile line is not observed in the case of the international profile.

Figure 2

Chilean Achievement Profile by Item Format
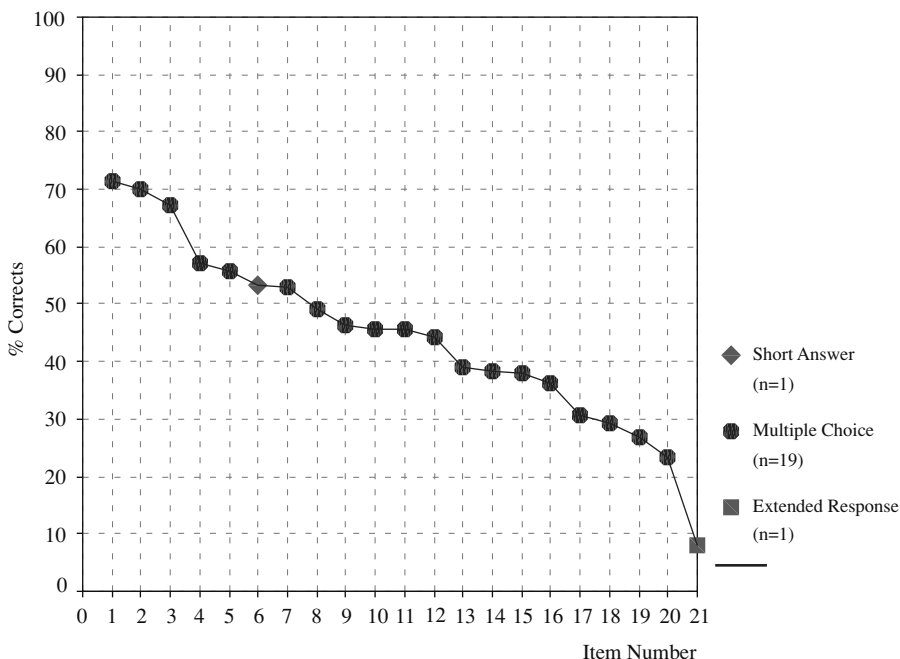
Source: TIMSS 1999 data files.

Considering that Item 21 is the only one with an extended response format, the possibility of a *country-format interaction* immediately arises. Even though Item 6 also has a free-response format, its overall structure is much more simple and straightforward than Item 21. In fact, the expected response in Item 6 consisted of just a number to be written over an answer line. In contrast, Item 21 has a much more complex structure. Item 21 is one of the released TIMSS items, and it is reproduced in figure 3.

A direct analysis of Item 21 allows not only an appreciation of its major format complexity, but also raises questions about its classification in the data representation content area. Which one is the most demanding task to solve this item: to understand the message of the two advertisements or to perform the operations required to solve the questions? If this second option were true, the item would be better classified in the fraction & number sense category.

As a final comment, it is also interesting to note that Item 21 was, not only for Chile but also for all the TIMSS countries, one of the most difficult questions in the test. In the international report, it was used to describe the type of problems that the top 10% students were typically able to solve (Mullis *et al.* 2000: 61).

While Item 21 describes what the most able students can do, Item 3 –the other released data representation item– describes what the students pertaining to the lower

Figure 3

Example of an Extended Response Item – Item 21

Chris plans to order 24 issues of a magazine. He reads the following advertisements for two magazines. *Ceds* are the units of currency in Chris' country.

Teen Life Magazine	Teen News Magazine
24 issues First four issues FREE The rest 3 *ceds* each.	24 issues First six issues FREE The rest 3.5 *ceds* each.

Which magazine is the least expensive for 24 issues? How much less expensive? Show your work.

Teen Life = 20
× 3
60 ceds
24 = 60 ceds

Teen News = 18
× 3.5
90
540
63.0 ceds
24 = 63 ceds

Teen Life is less expensive by 3 ceds.

quarter benchmarking can do (Mullis *et al.* 2000: 85). As shown in the next figure, Item 3 is a multiple-choice item that requires a straightforward lecture of data.
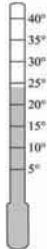
Chilean students obtained 67.40% of correct responses in this question, compared to 79.20% obtained by the international composite. The difference is somewhat smaller than the average for all items of the sub-scale (11.80 versus 17.25 percent points).

Figure 4

Example of a Multiple-Choice Item – Item 3

This table shows temperatures at various times on four days.

TEMPERATURE					
	6 a.m.	9 a.m.	Noon	3 p.m.	6 p.m.
Monday	15°	17°	24°	21°	16°
Tuesday	20°	16°	15°	10°	9°
Wednesday	8°	14°	16°	19°	15°
Thursday	8°	11°	19°	26°	20°

Thermometer

On which day and at what time was the temperature shown in the table the same as that shown on the thermometer?

A. Monday, Noon
B. Tuesday, 6 a.m.
C. Wednesday, 3 p.m.
D. Thursday, 3 p.m.

M022260

Testing the Performance Expectations Hypothesis. Do items aimed to measure the same cognitive skills cluster together in the achievement line? Figure 5 presents the profile of Chilean achievement in data representation items, according to their performance expectations.

As shown in the figure, using complex procedure items are spread out all over the distribution, ranging from Item 1 to Item 20. This is not surprising considering that more than half of the data representation items are part of this group. The other higher order set of items –solving & predicting– is somewhat more unified, with Items 21, 19, 18, and 17 forming part of a clear cluster with the lowest p values (0.31 and below). In a similar way, recalling & representing, as well as performing routine procedure items are somewhat more clustered toward the less difficult items, with p values 0.39 and upper.

From a developmental perspective, these results fit well the expectations: items aimed to measure higher order skills have lower p values, and items aimed to measure basic skills have higher p values. Taking from an educational perspective now, it is necessary to note that the Chilean curriculum at grade 8 –as well as the intended curriculum of most countries at this level– puts a major emphasis in the mastering of basic skills and the understanding of mathematical problems (Ministerio de Educación 1999). Since the intended curriculum reflects the steps of cognitive development, it is hard to disentangle a curricular effect from a developmental one.

Figure 5

Chilean Achievement Profile by Performance Expectation
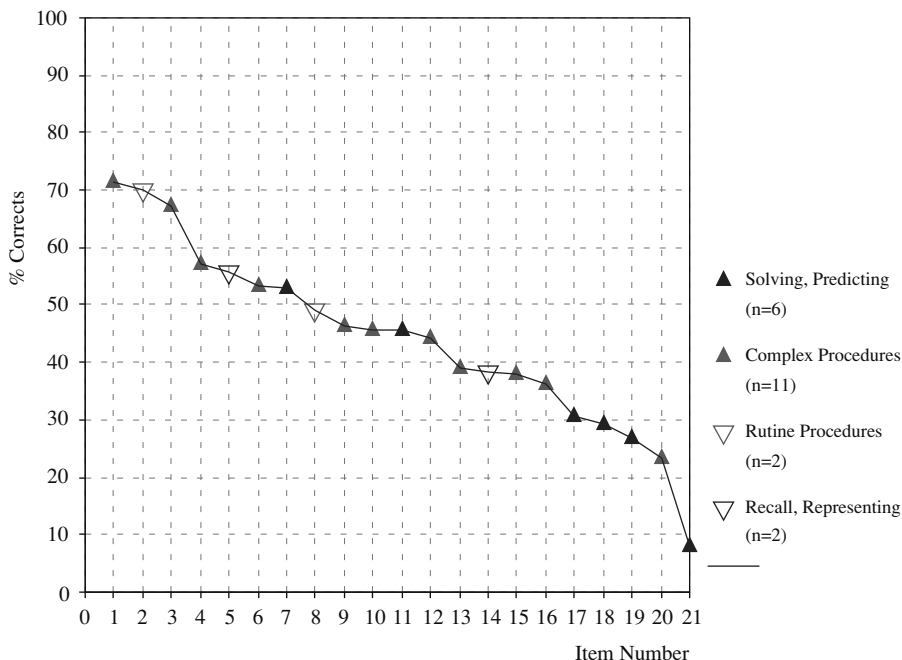
Source: TIMSS 1999 data files.

In spite of this emphasis on basic skills, the old Chilean curriculum covered a wide range of skills. Knowing, representing, using routine procedures, using more complex procedures, understanding mathematical problems, solving problems, and communicating were all mentioned in the framework; predicting was the only skill not intended to be taught (Ministerio de Educación 1999). Do the predicting items show a lower *p* value compared to the other data representation items? Since there are only two predicting items in the sub-scale (Items 7 and 11), it is hard to generalize. However, none of them is among the most difficult questions in this content area.

Textbooks can be understood as a link between intended and implemented curriculum. They serve as an interface between the national standards/objectives and the lessons put in practice by teachers at a classroom level. In Chile, textbooks are widely used by teachers, being for many the curriculum *de facto* when preparing classes and/or teaching. Assuming that the topics and skills presented in the textbooks have a greater chance to be taught to the students, it is interesting to know to which degree the TIMSS performance expectations were part of the math textbook used the year 8th grade students took the TIMSS test.

Consistent with the old curriculum, the Chilean textbook devoted extensive percent of its overall space to basic skills as knowing (41-50%) and using routine procedures

(31-40%)[8]. Higher order skills such as investigating, or problem solving & mathematical reasoning, were only given 11-20% and 1-10% of the textbook space, respectively (Ministerio de Educación, 1999). These emphases support the acquisition of basic skills, and seem to be reflected in the results shown in the previous figure.

Another focus of analysis is concerned with the number of items presented per cognitive skill. As shown in the previous figure, there is an uneven distribution of items per category. While in the two higher order skills there are six solving & predicting items and 11 using complex procedures items, in the two lower order skills there are just two items in recalling & representing and another two in performing routine procedures.

This unequal distribution of items can be justified by the higher order cognitive skill emphasis of TIMSS in the assessment of 8th grade students. However, the distribution of items per cognitive skills in the whole test presents a much more even distribution. This disparate allocation of items per skill in the data representation sector could be explaining in part the relatively better performance of Chilean students in this sub-scale as compared to the others.

Testing the Sub-Content Areas Hypothesis. The international report describes three different sub-content areas as constituting the data representation sub-scale (Mullis *et al.* 2000: 93):

–   Representation & interpretation of data in tables, charts, and graphs.
–   Means & ranges.
–   Simple probabilities – understanding & calculations.

In order to see if the items belonging to the same sub-content clustered together in the Chilean achievement profile, each one of the 21 data representation questions were classified according to these categories and then plotted in the figure 6 below.

As shown in the figure, the number of items varied considerably depending upon the topic under analysis. While in representation & interpretation of data there were 13 items, means and ranges only counted with one. Probability was just in the middle, with seven items addressing this topic.

Item 2 was the only means & ranges item, and it was among the easiest items in the subscale, with 70% correct responses. Representation & interpretation of data items are spread all over the range of achievement. In fact, both the easiest and the most difficult items are part of this sub-content category. Considering that more than half of the data representation items are aimed to measure representation & interpretation of data, it is not really surprising that there is not a special pattern in their distribution.

The probability items appear slightly clustered together and inclined towards the lower bound of the achievement line. This is especially evident for Items 17, 18, and 19. As we will see in the next section, this pattern can be easily linked to the topics intended to be taught in the old Chilean curriculum.

---

[8]   The overall textbook space is greater than 100%, given that a same block could be coded as stimulating two or more performance expectations at the same time. Details about the methodology used can be found in Schmidt, Raizen, Britton, Bianchi, & Wolfe (1997).

Figure 6

Chilean Achievement Profile by Content Sub-Area
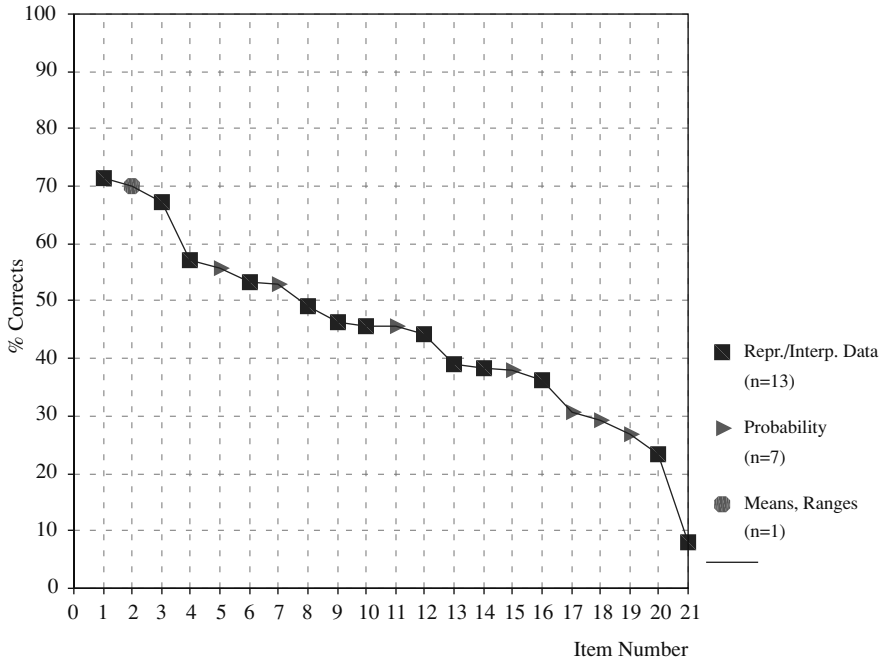
Source: TIMSS 1999 data files.

Testing the Curricular Intentions Hypothesis. Each item in the test was classified according to its status in the Chilean national curriculum. If the curriculum specified that students up to grade 8 should be taught the contents/skills necessaries to solve the item, the item was classified as intended. On the other hand, if students were not expected to be able to solve the problem according to the curriculum specifications, the item was labeled as not intended.

In order to see if intended items clustered together among the easiest questions in the subscale, the 21 data representation questions were plotted in the achievement profile line, marking them according to their curricular status (figure 7).

A clearer achievement pattern arises. According to expectations, intended items clustered together towards the easiest items, whereas not intended items clustered towards the most difficult ones. Intended items prevail from Item 1 to 14. Only three not intended items are counted among the half of easier items in the subscale: Items 5, 7, and 11.

Against expectations, the most difficult item of all (Item 21) is an intended one. Once more, this is the extended response item whose classification in the data representation sub-scale was questioned. Both factors taken together –item format and topic– could explain why this item stands so far apart from the remaining intended questions in the achievement profile line.

101

Figure 7

Chilean Achievement Profile by Curricular Intentions
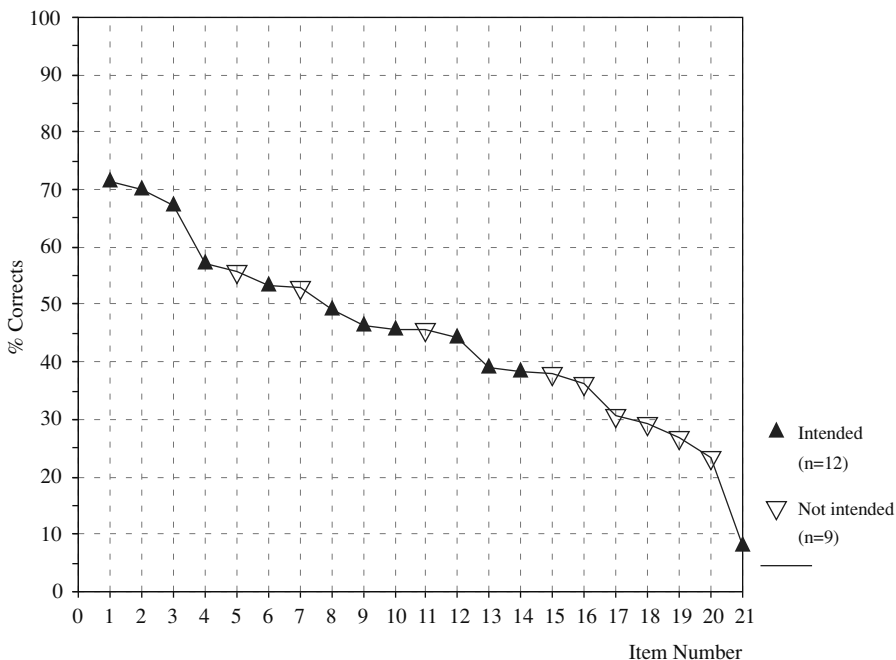
Source: TIMSS 1999 data files.

Turning now to the unintended items, they mostly range from Items 15 to 20, presenting a percent of correct responses below 39. Separated from this group stand Items 5, 7 and 11, presenting scores even above the Chilean data representation mean of 44%.

A comparison of Figures 6 and 7 reveals that not intended items include all the probability questions, plus two representation & interpretation of data items. Probability is, in fact, a topic only introduced at grade 12 in the old national curriculum. Nevertheless, representation & interpretation of data is a topic intended to be introduced from grade 6 throughout grade 8 (Ministerio de Educación 1999). Why were two items of this sub-content area (Items 16 and 20) classified as not intended? None of them required to predict –the only cognitive skill not intended in the curriculum. Even though it is always a possibility that a special type of graph or topic– beyond the curricular expectations for 8th grade students – was used in these items, this inconsistency raises questions about the validity and reliability of the data[9].

---

[9]   Items were marked as intended or not intended one by one, by a curricular specialist who used the Chilean framework as the criterion to make this decision.

Consistently with the curricular intentions, the remaining 11 representation & inter-pretation of data items are presented as intended in figure 7, as well as the only item aimed to measure means & ranges (Item 2). Means & ranges was expected to be intro-duced early in primarily education and was supposed to be reinforced throughout grade 8 (Ministerio de Educación 1999). These curricular intentions supported the high percent of correct responses (70%) given to Item 2.

In order to know if the differences in the percent of correct responses between intended and not intended items were significant, an independent sample t test of the mean was run. Since at this point there are reasonable explanations why Item 21 is so out of the pattern of the intended items, this question was excluded from the analysis. Intended items were correctly answered by almost 53% of the Chilean students, as compared to just 37,64% of correct responses obtained in the not intended items. These differences prove to be significant at $p < 0.01$. As we will see later, these results are somewhat opposed to the overall findings presented in the international report.

Testing the Implemented Curriculum Hypothesis. In figure 8, two separate percentages of correct responses were calculated: one for those students who were already taught the sub-content area addressed by each item, and another for students who were not yet

Figure 8

Chilean Achievement Profile by Curricular Implementation
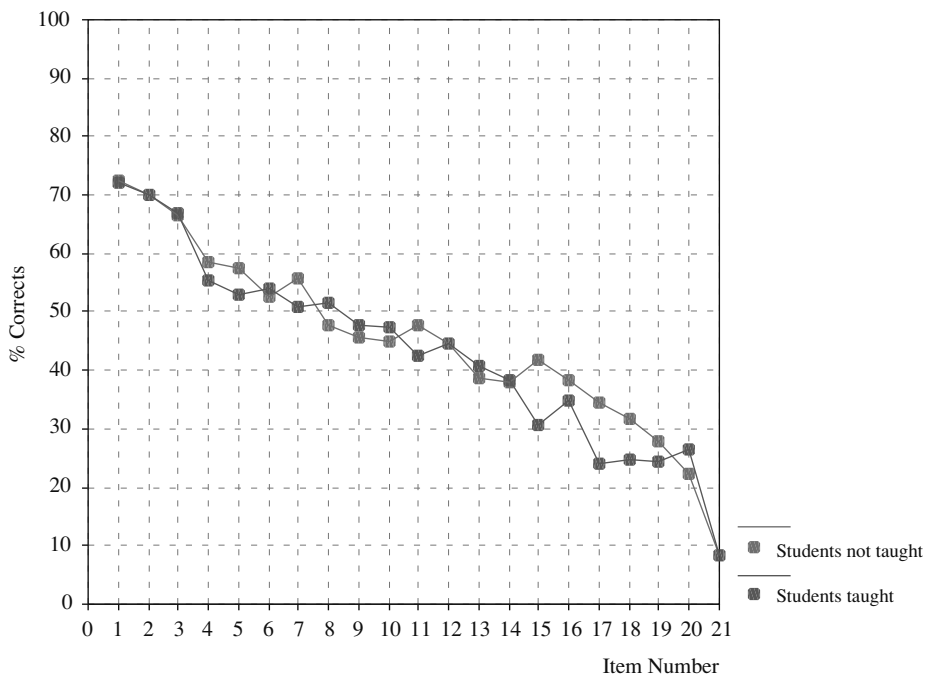
Source: TIMSS 1999 data files.

taught the sub-content area. The two groups were formed based on the teacher's report of each student. This procedure was followed in order to see if the percent correct for each item was higher for the first group (already taught) compared to the second (not yet taught).

Against expectations, the figure shows no substantial differences between students who were taught the topics versus those who were not. For both groups, the achievement profile lines follow a quite similar pattern, with slight variations from item to item. In the best of cases, these differences favored the students who were already taught the topic; this actually happened in 10 out of 21 items. In the remaining 11 items, eight graders not yet taught the topic did better than their counterparts already taught. Even thought these differences are not significant, on average, students not yet taught the topics obtained better results than those who were already taught (45.01% versus 43.26%, respectively).

Surprisingly, for all seven probability questions (Items 5, 7, 11, 15, 17, 18, and 19), students not yet taught the topic obtained higher percent of correct answers than those who were already taught the topic. These items also presented the seven widest differences in the percent of correct answers between both groups, ranging from 11 to 3.70 percent points in favor of the not yet taught group.

CONCLUSIONS AND DISCUSSION

This report presents a series of analyses aimed to foster understanding of the Chilean performance in the data representation sub-scale of the TIMSS 1999 assessment.

The findings suggest a *country-format* interaction. By far, the most difficult item was also the only extended response question (Item 21) included in the data representation sub-scale. Knowing that TIMSS 1999 was the first time that the Chilean eight graders were exposed to a standardized test with mathematics open-ended questions[10]; this interpretation makes even more sense. These findings are consistent with those reported by O'Leary (2001) when comparing the relative standing of countries across the three types of item format. He found that Irish students did relatively better on extended-response questions, which is consistent with the tradition of more open-ended essay type test used in this country. Thus, it seems that the frequent exposure to test format make a difference in student achievement.

If this holds true, a validity question concerning the Chilean relative performance in the five content areas arises. Is it by chance that algebra, the sub-scale with the biggest proportion of extended response items (7 out of 35), was also the content area with the weakest results for this country? In order to avoid this kind of uncertainty, it is recommended to keep more or less the same proportion of items with a different format across the test sub-scales.

The analysis of items by *cognitive skills* (performance expectations) revealed that higher order skills questions were somewhat clustered together towards the more difficult items, while the less demanding questions had lower difficulty levels. This pattern

---

[10]   The Chilean National Assessment System (SIMCE) included open-ended questions for the first time in the 8th grade math test of year 2000.

is consistent with cognitive development and with the cognitive skill emphasis of the old curricular framework.

Turning now to the *topic analysis*, the uneven distribution of items in each one of the three sub-content areas –representation & interpretation of data, ranges & means, and probabilities– make it hard to find a pattern in the achievement profile line. The labeling of three sub-content areas creates false expectations towards the number of items per topic. Clearly, to have just one item under the ranges & means category versus 13 under representation & interpretation of data does not seem to be a wise solution. A more proportionate distribution of items per topic would help to improve the quality of the test and the power of curricular interpretations.

The analysis of the released items raised concern about their classification in the different content categories. Strictly speaking, if Item 21 were to measure data representation skills, its questions should be focused on the information explicitly presented in the ads, and not in the inferences the students can make through mathematical operations.

Considering that the content area of fraction & number sense –together with algebra– presented the weakest results for Chile, one is lead to think that the presence of mathematical operations in Item 21 could be explaining, at least in part, the unexpectedly low *p* value of this question.

The analysis of items by *curricular intentions* provided interesting insights to better understand the profile of data representation achievement. According to expectations, Chilean students did significantly better in the data representation items intended to be taught up to grade 8 (once we exclude from the analysis the only extended response item of the sub-scale). These results are somewhat opposed to the findings of the international report, where no curricular effect is reported in the overall test.

In Chile, the overall percent of correct responses in math items across the entire test was 31%, whereas its average percent of correct responses in intended items was only 32% (Mullis *et al.* 2000: 350). A somewhat similar situation is observed when looking across the countries. The ranking calculated using all the items in the math test compared with the one obtained using only the intended items for each country did not show substantial differences in the relative position of the countries[11]. "It is clear that the selection of items does not have a major effect on the general relationship among the countries" (Mullis *et al.* 2000: 349).

A reasonable explanation for these discrepancies could be that, when we talk about different scores for intended and not intended items for Chile, we are really talking about differences between two ways of learning mathematics. Representation & interpretation of data is a more "common sense" topic, from which students can pick up the rudiments either in other curricular subjects or from the media, in particular newspaper and TV (Howson 2001). Probability is more strictly tied to school mathematics. Thus, the better performance of Chilean students in representation & interpretation of data may be better explained because of their "street" experiences with this subject, and not so much because of the instruction received in the schools.

---

[11]   It is worth noting that Chile was the country with the minor proportion of items (score points) identified as appropriate in the Curricular Matching Analysis procedure, with only 58% (98 score points) of the 162 items (169 score points).

Item 5 is a special case in this respect. Even though it is a probability topic, the problem presented is extremely familiar to 8th graders: the probability to toss a coin. Chilean students did unexpectedly well answering this question, equaling the international average. In the case of Chile, it seems we are more directly confronted with some kind of "street mathematics" phenomenon and less with a school related learning effect.

This idea is supported by the findings of the *implemented curricular* analysis, where no instructional effect was found in the achievement profile of Chilean students. Figure 8 is disappointing. It shows that achievement in the data representation sub-scale is not affected by instruction. Students who were already taught the topics did not do better than their counterparts who were not yet taught the contents; they did slightly worse. Are we confronted with evidence of the lack of effectiveness of the Chilean instruction, at least in the data representation area? Probably. At this point one wonders if something similar happens in other content areas and/or in other countries.

Finally, some comments about the *discrepancies between intended and implemented curriculum*. According to the old Chilean curriculum, probability topics were only supposed to be introduced at grade 12. Nevertheless, according to their teachers, 35% of eight graders were already taught this topic. On the other hand, representation & interpretation of data was an intended topic from grades 6 through 8, but only 49% of eight graders were already taught this topic. The same percent of students are reported to be taught arithmetic mean, a topic that should be taught well before grade 8 (Mullis *et al.* 2000: 179).

What kind of pedagogical guidance was providing the old curriculum? Why did teachers not follow the old framework? A plausible explanation is because it was an outdated document. Mathematics has sufficiently changed since 1980, the year this curricular framework was introduced. It is reasonable to think that teachers were not following it in order to give way to new math tendencies. In fact, these new tendencies were especially promoted in the context of the development of the new curriculum. Even though this new framework was ready since 1996, the new curriculum was not supposed to be introduced at grade 8 until the school year of 2002, when a new detailed program of study, derived from the framework, will be ready for use at the school level.

Thus, it is reasonable to think that, during these years, teachers have been involved in some kind of confusion and uncertainties about which curricular framework to use: the "old" or the "new". After the introduction of the new curriculum, it should be recommended to closely monitor its implementation in the Chilean classrooms. Maybe this could help in the promotion of better instruction and learning capabilities.

REFERENCES

GONZALEZ, E. J., J.A. MILES. (Eds.). (2001). *TIMSS 1999 User Guide for the International Database*. Chestnut Hill, MA: International Study Center Lynch School of Education Boston College.

HOWSON, G. (2001). *TIMSS, common sense and the curriculum*. Manuscript submitted for publication.

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.* 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]. Unpublished raw data.

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 Eighth Grade.* 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).* Paper presented at the 2001 annual meeting of the American Educational Research Association, Seattle, WA.

TIMSS 1999 International Database 1999 Assessment Data [CD-ROM]. (2000). Chestnut Hill, MA: International Study Center Lynch School of Education Boston College [Producer and Distributor].

SCHMIDT, W. H., S. A. RAIZEN, E. D. BRITTON, L. J. BIANCHI, R. G. WOLFE. (1997). *Many Visions, Many Aims Volume 2: A cross-national investigation of curricular intentions in school mathematics.* Dordrecht, The Netherlands: Kluwer Academic Publisher.