OECD publishing

Please cite this paper as:

Wu, M. (2010), "Comparing the Similarities and Differences of PISA 2003 and TIMSS", *OECD Education Working Papers*, No. 32, OECD Publishing.
http://dx.doi.org/10.1787/5km4psnm13nx-en

OECD Education Working Papers
No. 32

Comparing the Similarities and Differences of PISA 2003 and TIMSS

Margaret Wu

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

22-Apr-2010

English - Or. English

DIRECTORATE FOR EDUCATION

EDU/WKP(2010)5
Unclassified

Cancels & replaces the same document of 14 April 2010

COMPARING THE SIMILARITIES AND DIFFERENCES OF PISA 2003 AND TIMSS

OECD Education Working Paper No. 32

This paper was authored by Margaret Wu of the Assessment Research Centre, University of Melbourne.

Contact:
Pablo Zoido; Email: pablo.zoido@oecd.org; Tel: +33 1 45 24 96 07

JT03282237

English - Or. English

OECD DIRECTORATE FOR EDUCATION

OECD EDUCATION WORKING PAPERS SERIES

This series is designed to make available to a wider readership selected studies drawing on the work of the OECD Directorate for Education. Authorship is usually collective, but principal writers are named. The papers are generally available only in their original language (English or French) with a short summary available in the other.

Comment on the series is welcome, and should be sent to either edu.contact@oecd.org or the Directorate for Education, 2, rue André Pascal, 75775 Paris CEDEX 16, France.

---

www.oecd.org/edu/workingpapers

ABSTRACT

This paper makes an in-depth comparison of the PISA (OECD) and TIMSS (IEA) mathematics assessments conducted in 2003. First, a comparison of survey methodologies is presented, followed by an examination of the mathematics frameworks in the two studies. The methodologies and the frameworks in the two studies form the basis for providing explanations for the observed differences in PISA and TIMSS results. At the country level, it appears that Western countries perform relatively better in PISA as compared to their performance in TIMSS. In contrast, Asian and Eastern European countries tend to do better in TIMSS than in PISA. This paper goes beyond making mere conjectures about the observed differences in results between PISA and TIMSS. The paper provides supporting evidence through the use of regression analyses to explain the differences. The analyses showed that performance differences at the country level can be attributed to the content balance of the two tests, as well as the sampling definitions – age-based and grade-based – in PISA and TIMSS respectively. Apart from mathematics achievement, the paper also compares results from the two studies on measures of self-confidence in mathematics. Gender differences are also examined in the light of contrasting results from the two studies. Overall, the paper provides a comprehensive comparison between PISA and TIMSS, and, in doing so, it throws some light on the interpretation of results of large-scale surveys more generally.

RESUME

Le présent document établit une comparaison détaillée des évaluations des mathématiques PISA (OCDE) et TIMSS (IEA), toutes deux menées en 2003. Il présente tout d'abord une comparaison des méthodologies d'enquête, puis un examen des cadres d'évaluation des mathématiques. C'est en effet par une analyse des méthodologies et cadres d'évaluation des deux études que l'on peut expliquer les différences constatées dans les résultats du PISA et ceux du TIMSS. Au niveau des pays, il apparaît que les nations occidentales réussissent relativement mieux à l'enquête PISA qu'à l'enquête TIMSS. En revanche, les pays d'Asie et d'Europe de l'Est ont tendance à obtenir de meilleures performances aux évaluations TIMSS qu'aux évaluations PISA. Au-delà de simples conjectures sur les différences observées, le présent document fournit des éléments de preuves en utilisant des analyses de régression qui expliquent les disparités de résultats. Les analyses ont démontré que les variations de performance au niveau national peuvent être imputées à l'équilibre des contenus des deux tests, ainsi qu'aux définitions d'échantillonnage – fondées sur l'âge ou fondées sur la classe – pour PISA et pour TIMSS, respectivement. Outre les performances en mathématiques, le présent document compare les résultats des deux études sur les mesures de la confiance en soi en mathématiques. Les différences entre les sexes sont également examinées à la lumière des résultats contrastés des deux enquêtes. Dans l'ensemble, ce document offre une comparaison complète entre PISA et TIMSS et, ce faisant, éclaire plus généralement les interprétations des résultats d'enquêtes de grande envergure.

FOREWARD

This paper was authored by Margaret Wu of the Assessment Research Centre, University of Melbourne, with contributions from Professor Kaye Stacey, University of Melbourne, and her team in classifying PISA and TIMSS items, and in providing references and comments for the manuscript.

The author would like to express her thanks to Professor Patrick Griffin for reading the manuscript and for providing comments.

TABLE OF CONTENTS

CHAPTER 1 - INTRODUCTION 7
Purpose of the working paper 7
Surveys for comparison 8
Similarities and differences 8
Focus of comparison 8
Background information about the two surveys 9
The aims of TIMSS and PISA 9
Organisation of the report 10
CHAPTER 2 - SURVEY METHODOLOGIES 12
Introduction 12
Population definition 12
Implications of the different definitions 13
Comparison of age distributions 14
Comparison of grade distributions 15
Sampling methods 17
Test characteristics 18
Test length 18
Test design 18
Amount of assessment material 19
Scaling methodology 20
Field operation procedures 22
Calculator use 22
Questionnaires 23
Summary 25
CHAPTER 3 - COMPARISON OF PISA AND TIMSS MATHEMATICS FRAMEWORKS AND ITEM FEATURES 28
Approaches to the development of the mathematics assessment frameworks 28
TIMSS mathematics assessment framework 28
TIMSS mathematics content domains 29
TIMSS mathematics cognitive domains 31
PISA mathematics assessment framework 32
Situation/context dimension 34
PISA mathematical content dimension – The overarching ideas 34
Mathematical processes dimension 37
Classifying items according to competency clusters 39
A Comparison of PISA and TIMSS Mathematics Frameworks 39
A comparison of PISA's content dimension with TIMSS' content dimension 39
exchange rate 40
Question 1: EXCHANGE RATE 40
staircase 41

Question 1: STAIRCASE 41
CARPENTER 42
Question 1: CARPENTER 42
internet Relay chat 43
Question 1: INTERNET RELAY CHAT 43
Question 2: INTERNET RELAY CHAT 43
exports 44
Question 2: EXPORTS 44
A comparison of PISA's processes dimension with TIMSS' cognitive domains 45
Characteristics of tests and items 46
Item Format 46
Amount of reading 47
Summary 48
CHAPTER 4 - COMPARISON OF PISA AND TIMSS ACHIEVEMENT RESULTS 50
Comparisons of country mean scores 50
Explaining the Differences between PISA and TIMSS Results 56
Years of Schooling and Age at time of testing 56
Impact of Years of Schooling on Student Performance in Mathematics 60
The impact of differences in content balance between PISA and TIMSS 64
Achievement by content domains 64
Predicting PISA Mathematics Country Mean Scores 68
Implications of differential performance of countries in content domains 71
TIMSS Data domain and PISA Uncertainty overarching idea 71
Examining the spread of PISA and TIMSS achievement distributions 72
Standard Deviations 72
Percentiles 74
The impact of calculator availability on differences in performance in TIMSS and PISA 77
Summary 77
CHAPTER 5 - GENDER DIFFERENCE AND ATTITUDES 79
Introduction 79
Gender differences 79
Gender differences for overall mathematics scale 79
Gender differences by mathematics content areas 82
Possible explanations for observed differences between PISA and TIMSS in gender gap 83
Gender difference in the spread of achievement distributions 84
Attitudinal scales 88
Self-confidence index 88
Interest and motivation indices 89
Gender differences in attitudes towards mathematics 91
Socio-economic Background of Students 93
Summary 94
CHAPTER 6 - CONCLUSIONS 95
Overview 95
The impact of content balance on achievement results 95
The impact of reading load on mathematics achievement results 97
Grade-based and age-based samples 97
Correlated factors 98
Gender and attitudinal differences 98
And finally… 99
References 100

CHAPTER 1 - INTRODUCTION

Purpose of the working paper

1. What would help explain differences between the results in students' mathematics performance from two respected international education surveys conducted in 2003: the OECD's Programme for International Student Assessment (PISA) and the International Association for the Evaluation of Educational Achievement (IEA)'s Trends in International Mathematics and Science Study (TIMSS)? This was a question asked by the PISA Governing Board (PGB) and was at the heart of the development of this report.

2. The results of these two international assessments have often been cross-referenced and synthesised to present an overall picture of mathematical achievements. Valid cross-references of the results from PISA and TIMSS require, however, a clear and accurate understanding of the two assessments in terms of their objectives, assessment frameworks and the nature of the tasks and items that were presented to students. The objective of this report is to provide such an understanding.[1]

3. This objective is important as comparisons are inevitably made between the results when two international surveys are carried out concurrently to measure mathematics achievement, and superficial comparisons can often be misleading, inaccurate or simply inadequate. For instance, it is customary to brush aside any comparison of PISA and TIMSS by stating that (1) PISA samples are age-based while TIMSS samples are grade-based; and (2) PISA is not curriculum-driven while TIMSS is based on curriculum. While these statements are generally correct, these explanations of the differences between the surveys hardly convey any useful information in gaining an understanding of the comparative results of the two surveys. Further, uninformed comparisons can be dangerous, particularly when somewhat emotionally charged. Prais (2003) made a number of conjectures in a comparison of PISA and TIMSS results for the United Kingdom. In a rejoinder to Prais' article, Adams (2003) pointed out that Prais' criticisms were based on some misunderstanding of the two surveys. Furthermore, a comparison limited to the results of one country is unlikely to have the power to reveal patterns of similarities and differences. As a result, this working paper attempts to provide a comprehensive comparison of PISA and TIMSS, examining both methodological similarities and differences, as well as similarities and differences in the results. Aspects in relation to methodology include sampling, framework development and test construction. Aspects in relation to results include comparisons and explanations of rank ordering of countries, observed gender differences, as well as the impact of attitude on achievement.

4. The major target audience of this report will be educational practitioners. The term "educational practitioners" is widely defined here to include policy makers and researchers in the education field. The report aims to help those involved in planning instructions and monitoring achievement in mathematics to understand the differences and similarities between PISA and TIMSS, and to provide guidance on how to

1 . This purpose was stated in the OECD's International Call for Tender for the development of thematic reports on PISA 2003 (OECD/EXD/PCM/EDU(2003)56, p. 4).

relate and interpret the results from the two assessments. For example, PISA and TIMSS results can identify relative strengths and weaknesses of students in the field of mathematics. This information, when interpreted correctly, can in turn be used for an evaluation of current practices as well as future reforms in curriculum and instruction.

Surveys for comparison

5. This report uses results from the PISA 2003 mathematics assessment and TIMSS 2003 population 2 mathematics assessment[2] as the basis for comparison.

6. There are two reasons for the choice of these two surveys for comparison. First, they were both conducted at around the same time[3]. The data collected represent a cross-sectional profile of students' mathematics achievement in 2003 in each participating country. On-going changes in educational reforms within each country are not likely to account for the differences between the results of the two surveys, as both surveys were conducted at about the same time. Second, although PISA assesses reading, science and mathematics, the majority of testing time was devoted to mathematics in PISA 2003 so that the data collected covered most mathematics content areas, as was typically the case in TIMSS studies. This makes the results of the two surveys more comparable.

Similarities and differences

7. While it is useful to identify differences between PISA and TIMSS, the identification of similarities between the two surveys should also be valuable. First, where the findings from both surveys agree, the results provide strong evidence for policy makers and researchers to take appropriate actions based on the findings. For example, if gender differences in mathematics performance from both surveys are consistent, then there is a clear message about the differential performance of girls and boys, despite differences in grade and age in the two surveys. Second, the identification of the extent of similarities in the survey methodologies could inform policy decisions regarding the best way to move forward so that the two surveys complement, rather than duplicate, each other. For example, both surveys are currently paper-and-pencil based tests with short tasks. It is possible that some extended performance tasks delivered through computer-based testing could be included in one survey to tap into a different aspect of mathematics performance, so that the two surveys can provide complementary information about all aspects of mathematics literacy. Consequently, a useful comparison between the two surveys should go beyond just looking for differences. This working paper provides a holistic comparison between the two surveys, identifying both similarities and differences between the methodologies and the findings.

Focus of comparison

8. More specifically, this report addresses four aspects of PISA and TIMSS:

- What findings are similar between the two surveys? Agreements between findings from the two surveys will reinforce the underlying messages and provide some evidence of validity of the results.
- What findings are different, or even contradictory, between the two surveys? What are possible explanations for the differences?

2. TIMSS Population 2 refers to the Grade 8 cohort. TIMSS Population 1 refers to the Grade 4 cohort.

3. The PISA testing window was between March and August 2003 (OECD, 2005, p.46). In TIMSS, seven Southern Hemisphere countries tested in October through December, 2002. Korea tested later in 2003. The remaining countries tested mostly between April and June 2003 (IEA, 2003, p.18).

- Which issues are investigated by only one survey? What findings from the two surveys are complementary?

- What lessons can we learn from a cross-comparison of the two surveys? How can we improve each survey based on the findings of this report?

Background information about the two surveys

9. PISA is conducted by the OECD while TIMSS is conducted by the IEA. To fully understand the differences between PISA and TIMSS, it will be important to be familiar with the history and the "philosophies" of IEA studies and OECD work. OECD, being a co-operative organisation between governments, has policy-makers' interests as the focus of its work. In contrast, IEA, formed as a united body of research organisations, has the interests of researchers at the forefront of its studies. Although the distinction between policy focus and research focus is a blurred one, as many research questions that drive TIMSS are heavily influenced by policy considerations. Nevertheless, the different backgrounds of the two organisations have resulted in setting different goals for the studies conducted. For example, in TIMSS it was deemed important to link the survey results directly to instructional practices in the classrooms, while in PISA, the measure of the outcome of schooling is deemed more important for governments in shaping educational policies. The emphases in the main goals of each study in turn had an impact on population definition and sampling procedures. For example, to examine instructional practices and relate these to student achievement, one needs to sample classes. To sample classes, the population definition will need to be grade based. In contrast, to compare outcomes of schooling, an age-based sample may place countries on more equal footings for describing the preparedness of students for adult life. Consequently, a clear understanding of the different objectives of each study is fundamental in subsequent analyses of the comparisons of the surveys.

The aims of TIMSS and PISA

10. In the introduction to the *TIMSS 2003 International Mathematics Report* (IEA, 2003), the aim of the study is stated as follows:

> The aim of TIMSS … is to improve the teaching and learning of mathematics and science by providing data about students' achievement in relation to different types of curricula, instructional practices, and school environments. (p.13)

11. The aim of TIMSS places teaching and learning at the forefront, with a special mention of linking achievement to curricula and instructional practices. In contrast, in the OECD publication *Learning for Tomorrow's World – First Results from PISA 2003* (OECD, 2004), the aim of PISA is stated as follows:

> PISA seeks to assess how well 15-year-olds are prepared for life's challenges. … focusing on young people's ability to use their knowledge and skills to meet real-life challenges, rather than merely on the extent to which they have mastered a specific school curriculum. (p.20)

12. PISA's statement of aim indicates that the link between achievement and curricula is not regarded as the main objective of the study. PISA adopts a "literacy" concept about the extent to which students can apply knowledge and skills. An assessment of this literacy in various subject domains will have direct policy relevance for governments. This is not to say that TIMSS does not produce useful information for policy-makers, or that PISA results do not inform teaching and learning. Such a division between policy focus and research focus is an oversimplification of the aims of the surveys. Many research objectives in TIMSS were driven by policy considerations, and many policy objectives in PISA result in

research themes. However, it is the relative emphases of the two studies that are different. When a study is designed for a main purpose, the results are usually not as readily useful for other purposes.

13. It is not surprising that PISA has a policy orientation while TIMSS has a research orientation, since the governing body of PISA consists of governmental departments, while institutional members of IEA are research centres that may or may not be linked to the government of each country. Consequently, the decision making processes in the two studies differ to some extent, since the participants of decision making meetings are not the same group of people. However, there is some overlap of participants. Around eleven countries have the same government department taking charge of both PISA and TIMSS in participating in the decision-making processes at the international level, and about half of these countries have the same person in charge of both projects[4]. From this point of view, PISA and TIMSS are not entirely separate studies, since common experiences and problems in these two surveys have often been cross-referenced when setting directions for each survey.

14. Nevertheless, the total number of countries participating in each study (41 countries in PISA 2003, and 50 countries[5] in TIMSS 2003 Grade 8 cohort) is much larger than the number of countries taking part in both studies (22 countries), and the cohort of countries that participated in each study has an impact on the directions taken for the development of the assessment instruments. First, in both studies, there has been an active involvement from participating countries in shaping the assessment instruments through contributions and reviews of items. Second, the target difficulty level and cultural balance also need to match the group of participating countries. Characteristics of items may be influenced by the background of the participants of the surveys. For example, specific item contexts may be selected or avoided depending on the level of socio-economic status of students in participating countries, since students' familiarity with the contexts can have some impact on the results. The items selected also need to work well in all countries and languages.

Organisation of the report

15. The content of this report is organised in six chapters.

16. Chapter 2 compares survey methodologies used in the two surveys. The topics considered include population definitions, sampling methods, test characteristics, scaling methods, field operations and questionnaires administration.

17. Chapter 3 provides a detailed comparison of the mathematics frameworks and test specifications used in PISA and TIMSS. An attempt is made to align the two frameworks according to both the content domains and cognitive domains.

18. Chapter 4 examines the degree of similarities and differences between TIMSS and PISA achievement results in terms of country mean scores. Various hypotheses are tested to explain the observed differences in country performance in TIMSS and PISA. With the identification of a number of factors, quantitative models are built to explain achievement differences.

19. Chapter 5 focuses on comparisons of gender differences and students' attitudes towards mathematics, as reported in TIMSS and PISA. The choice of these student level variables was based on the availability of published results provided by the two surveys.

4. It is difficult to obtain these figures precisely, since there are often changes of personnel or changes of contractors for running each study.

5 Some of the participants in TIMSS are sub-regions of a country, for example, Basque country in Spain and two provinces of Canada.

20. Chapter 6 summarises the findings, as well as draws attention to the implications of the findings for PISA and TIMSS, and, in fact, for large-scale assessments more generally.

CHAPTER 2 - SURVEY METHODOLOGIES

Introduction

21. This chapter compares survey methodologies used in PISA and TIMSS. In particular, five aspects of survey methodologies are compared: sampling, test characteristics, scaling methods, field operations and test administration. On the whole, both PISA and TIMSS adopt similar survey methodologies typically used for large-scale studies. Both surveys chose the methodologies to meet survey objectives while taking account of the constraints.

22. As international studies, both surveys attempt to provide information on mathematics achievement at the national level for participating countries. Since it would be impractical to test every student in each country, both PISA and TIMSS use sampling methodology to select a representative sample from each country. The use of samples leads to the implementation of a set of statistical procedures appropriate for drawing inferences from samples about the population.

23. Both PISA and TIMSS have world experts in large-scale survey sampling methodology providing clear directions for sampling within each country. The sampling methodologies used in both studies are similar.

24. In addition, both surveys use similar methodologies for estimating student performance and the construction of a proficiency scale. These scaling methodologies are described in more detail below. There is also some overlap in expertise in this scaling process, as some members of the technical advisory group for PISA have also been involved in the scaling of TIMSS and the United States National Assessment of Educational Progress (NAEP) [6] data.

25. In terms of field operations, PISA and TIMSS have similar processes, with minor variations in translation and verification procedures. As technical aspects of methodologies for international surveys are described in detail in published reports, there is now a move towards forming standards for conducting international, or large-scale, surveys. Consequently, it is not surprising that survey methodologies are becoming more similar across different studies in broad terms, but there are variations at the level of details.

26. The following sections compare each aspect of survey methodologies between PISA and TIMSS in more detail.

Population definition

27. PISA and TIMSS have different population definitions. The target population of PISA is defined as follows:

> The desired base PISA target population in each country consisted of 15-year-old students attending educational institutions located within the country, in grades 7 and higher. (p.46, OECD, 2005)

6. TIMSS uses essentially the same scaling methodology as NAEP.

28. Note that while PISA has an age-based population definition, there is a reference to grade so that only students in grade 7 and above are included in the sample.

29. TIMSS Grade 8 population definition for the 2003 survey is the following:

> All students enrolled in the upper of the two adjacent grades that contain the largest proportion of 13-year-olds at the time of testing. This grade level was intended to represent eight years of schooling, counting from the first year of primary or elementary schooling, and was the eighth grade in most countries. (p.110, IEA, 2004)

30. The fact that TIMSS results are labelled as "Grade 4" or "Grade 8" has misled some to think that Grade 8 students are selected from every country. The term "Grade 8" refers to eight years of schooling, rather than a grade labelled as Grade 8 in every country. Therefore, the TIMSS population definition aims to control for the number of years of schooling. Note that while TIMSS has a grade-based sample, in fact, there is a clear reference to age in selecting the appropriate grade for testing. That is, while the sample is grade-based, the target population aims to capture 13 or 14-year-olds. This additional reference to age provides some guidance to countries to select the correct grade. It helps to overcome some difficulties in defining "the first year of primary schooling", as there are variations across countries in formal and informal education for very young children. In this sense, the desired population base of TIMSS is somewhat age-based, but the operational population definition is grade-based.

Implications of the different definitions

31. What are the implications of the differences in population definitions in PISA and TIMSS? In broad terms, PISA examines the educational yield in the first 15 years of life of a child, and TIMSS examines the educational yield of the first eight school grades in each country.

32. So one might say that PISA asks the question: "In each country, what has the education system been able to do to raise the mathematics achievement of a child by the time he/she reaches 15 years of age?" In contrast, TIMSS asks the question: "What has eight years of school grades achieved in raising the mathematics standard of a child?" In PISA, if Country X had lower mean achievement than Country Y, one might conclude, in simplistic terms, that Country X had not been able provide as much effective education for a child in the first 15 years of his/her life. In TIMSS, if Country X had lower mean achievement than Country Y, one might conclude, in simplistic terms, that the first eight grades of schools in Country X had not been as effective as the first eight grades of schools in Country Y. Of course there are many other variables (such as socio-economic status of students or culture traditions that are not easily amenable to policy manipulation) that need to be taken into account to make conclusions about the effectiveness of any educational system, and the pictures are usually not as simplistic as the above examples illustrate. See Box 2.4 for further information on student background variables selected for providing insight into the making of effective education systems.

33. The population definitions of PISA and TIMSS are consistent with the objectives of each survey. If one wants to evaluate the effectiveness of the provision of an education system for each child when he/she reaches the age at which he/she can leave school, then the PISA population definition will provide a more comparable sample. On the other hand, if one wants to evaluate the effectiveness of instructional practices in classrooms, then the TIMSS population definition will provide a more comparable sample as it controls for the number of school grades a child attends. In some sense, PISA takes a look at a bigger picture of education systems as a whole. In contrast, TIMSS focuses on specific issues of education, namely, schools and classrooms.

Comparison of age distributions

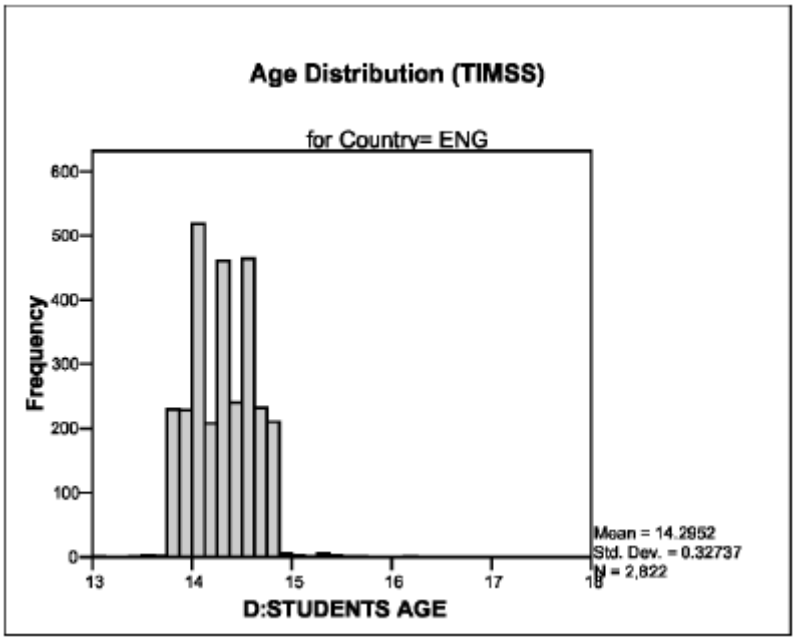
34. The international average age of students in TIMSS "Grade 8" assessment was 14.5, with country mean age ranging from 13.7 (Scotland) to 15.5 (Ghana) (See Exhibit 2, IEA, 2003, pp. 20-22), while the international average age of PISA 2003 students was 15.8[7], with country mean age ranging from 15.7 to 15.9. Of course, since PISA samples are selected by age, there is little variation in student age across countries.

35. Within each country, there is also a spread of ages. As expected, in TIMSS, the variation of age within each country is greater than the variation of age in PISA. The following provides a summary of the comparison of age distributions between PISA and TIMSS.

36. Students in PISA typically are aged between 15 years and 3 months and 16 years and 2 months, within each country. That is, there is a one-year age span within each country's sample. The spread of students' age distribution is similar across all countries.

37. In TIMSS, there is typically a two-year age span in each country's student sample. However, there are differences across countries. For example, in England, the age difference between students is generally less than one-year, as shown in the histogram in Figure 2.1.

Figure 2.1 Age distribution of students in the TIMSS sample for England

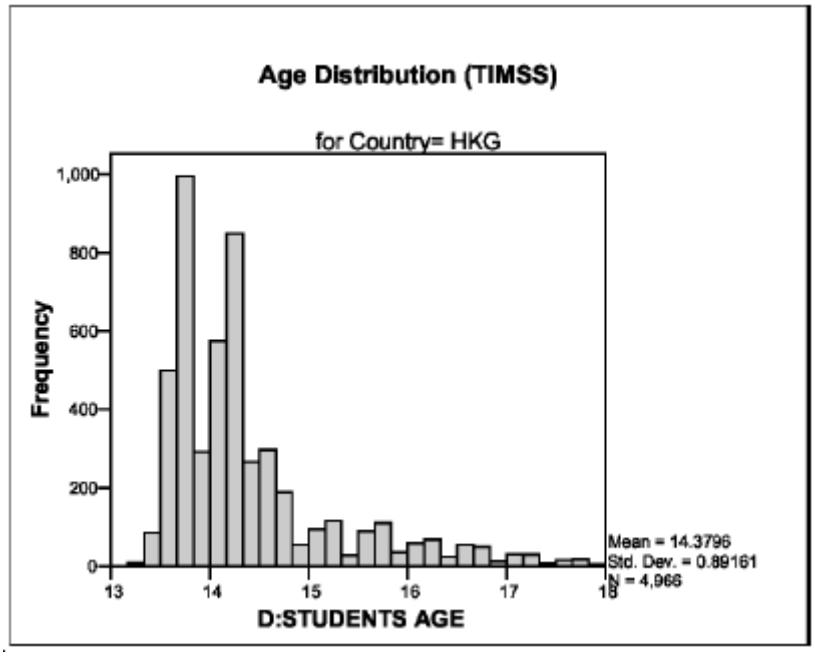


38. That is, both the TIMSS sample and PISA sample for England have similar spread of age distribution, with a range of around one year.

39. In contrast, for Hong Kong-China, the age distribution for the TIMSS sample has a large spread, as shown in Figure 2.2.

7. Computed across all countries in PISA 2003, with each country contributing equal weight.

Figure 2.2 Age distribution of students in the TIMSS sample for Hong Kong-China



40. That is, the age of students in the TIMSS sample for Hong Kong-China covers around four years, while the age differences of students in the Hong Kong-China PISA sample are within one year (but across a number of grades).

41. The United Kingdom and Hong Kong-China examples are two extreme cases. For most countries, the age distribution of TIMSS samples covers around two years.

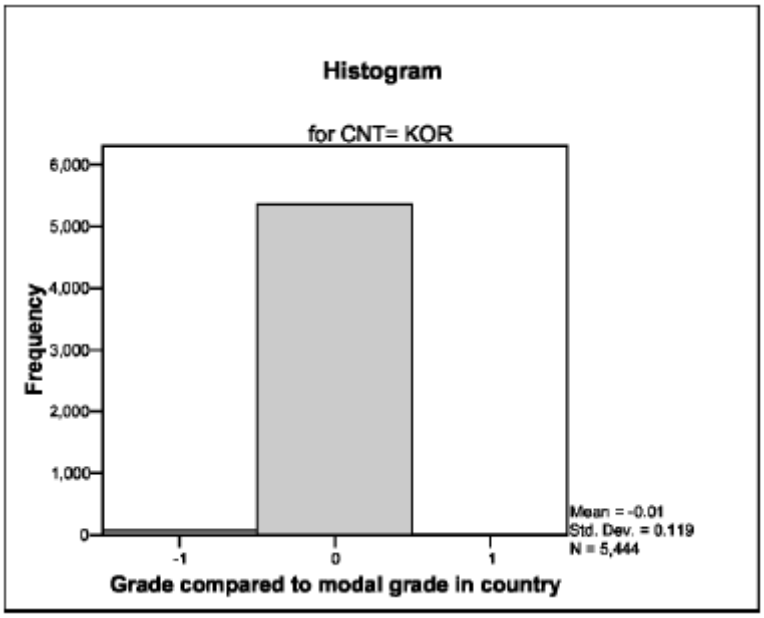
42. The age distributions of sampled students may have an impact on student achievement. For example, in TIMSS, different countries have different age cohorts due to country-specific regulations such as age of entry into schools or policies on retention. The upper of the two adjacent grades that contain the largest proportion of 13-year-olds could have more 14-year-olds than 13-year-olds, or have only 13-year-olds. A country such as Norway, where children enter schools at a relatively younger age, may have younger students at Grade 8, compared with countries where students enter schools at an older age.

43. In contrast, the PISA sample could draw students from predominantly one grade level, or from two or more grades, depending on the distribution of 15-year-olds across grades. The implications of multiple grades versus single grade will need further investigation. The following shows some examples of grade distributions in PISA.

Comparison of grade distributions

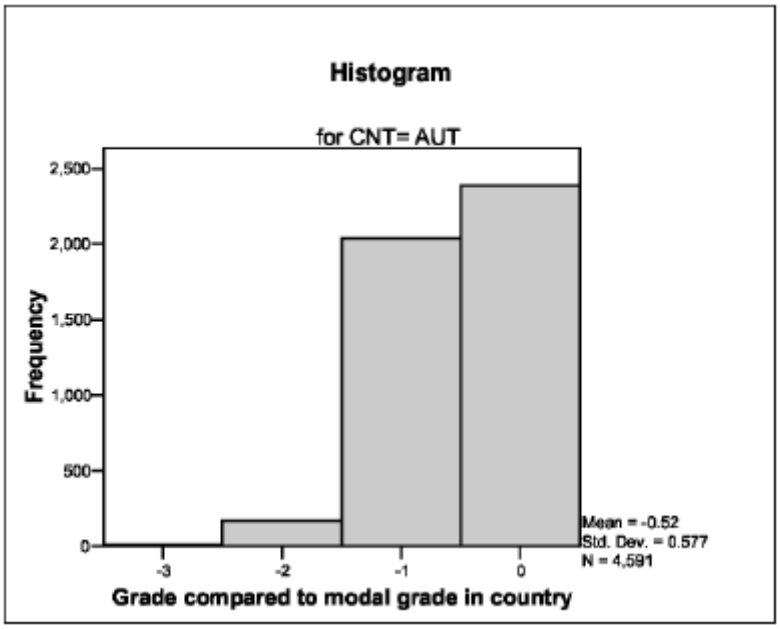
44. Since TIMSS samples are grade-based, all students in each country's sample are in the same grade. In contrast, for PISA, 15-year-olds may be in a number of different grades in each country. There are considerable variations across countries in terms of the number of different grades covered in the PISA sample. For example, in Iceland and Japan, all students in the PISA sample are from the same grade. In Korea, the majority of students are from the same grade, as shown in Figure 2.3.

Figure 2.3 Grade distribution of PISA sample in Korea



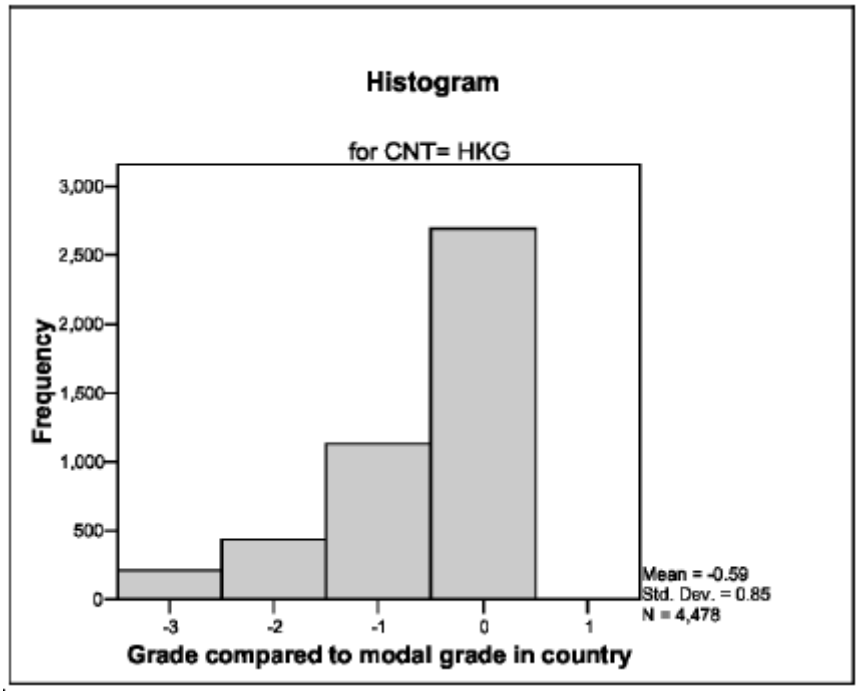
45. In some countries, the majority of the students in the PISA sample come from two grades, for example, in Austria and the Czech Republic (see Figure 2.4, for an example).

Figure 2.4 Grade distribution of PISA sample in Austria



46. In some countries, there are many 15-year-olds who are in grades below the modal grade, such as in Hong Kong-China and Portugal.

Figure 2.5 Grade distribution of PISA sample in Hong Kong-China



47. One might hypothesise that, if many 15-year-olds in a country are in grades well below the modal grade, then an age-based sample will tend to lower the average performance of students compared to a grade-based sample. On the other hand, in countries where there are many 15-year-olds who are in higher grades than the modal grade, one might expect 15-year-olds to perform relatively better than just students in a single grade, since school grade level reflects the number of years of schooling. In comparing country performance in PISA and TIMSS, the grade distribution of the PISA sample in each country, and the age distributions of the TIMSS samples, should be taken into account. These are discussed in more detail in Chapter 4.

Sampling methods

48. The sampling design for both PISA and TIMSS is a two-stage stratified sample for most countries. The first stage of sampling is the selection of schools from a list of all schools satisfying the target population definition for the survey. The selection of schools is carried out using probability-proportional-to-size (PPS) method. That is, the probability that a particular school will be selected is proportional to the number of eligible students in that school. When an equal number of students are then selected from each of the chosen schools, the probability that a particular student will be chosen will be the same for all students in the population. If PPS is not used, and each school has the same probability of being chosen, then students from very small schools are more likely to be in the sample than students from very large schools. Consequently, there will likely be larger standard errors and bias of estimates if PPS is not used.

In PISA, the second stage of sampling is the selection of, typically, 35 students from each sampled school, while, in TIMSS, the second stage of sampling is the selection of one intact class[8]. In PISA and TIMSS, various standards have been set in relation to school level and student level exclusions, and both surveys

8. In most countries, one class per school was selected. But some countries selected two classes per school.

attempt to minimise the proportion of exclusions to ensure that the sample collected is representative of the target population for each country.

49. Both PISA and TIMSS require a minimum of 150 schools to be selected in each country. PISA also recommends that the total number of sampled students is at least 5250, while TIMSS requires a minimum of 4000 students to be selected in each country.

50. The achieved degree of precision for estimates of mean student performance by country is not too different in magnitude in PISA and TIMSS, with standard errors slightly smaller in PISA than in TIMSS (see Table 9 in Chapter 4). For the 22 countries that participated in both PISA and TIMSS, the average standard error for country means is 3.2 score points in PISA, and 3.3 score points in TIMSS, although the score points are not on the same scale. If PISA scores are placed on the TIMSS scale (with the same overall TIMSS standard deviation for the 22 countries), then the average standard error (over the 22 countries) in PISA is 2.7 score points on the TIMSS scale, about 0.6 score points less than the average standard error in TIMSS.

Test characteristics

Test length

51. In PISA, each student took two hours of testing, with a break after one hour. In TIMSS, the testing time for each student was 90 minutes and was divided into two sessions, with a break in between.

Test design

52. To allow for coverage of all mathematics content areas in the assessment, but at the same time not placing too much burden on individual students, both PISA and TIMSS utilise matrix sampling test design (see Box 2.1). That is, test items are placed in a number of test booklets with linking items across booklets, and each student takes only one test booklet. In this way, each student only answers a fraction of the test items, out of a large number of items being tested in the whole assessment.

53. In PISA, 13 test booklets were rotated among students, with each booklet containing items from at least two subject domains from mathematics, reading, science and problem solving. In addition, every booklet contained at least one mathematics item cluster. In TIMSS, 12 test booklets were rotated among students, with each booklet containing both mathematics and science items.

Box 2.1 Matrix sampling test design in PISA and TIMSS

Cluster rotation design used to form test booklets for PISA 2003

Booklet	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	M1	M2	M4	R1
2	M2	M3	M5	R2
3	M3	M4	M6	PS1
4	M4	M5	M7	PS2
5	M5	M6	S1	M1
6	M6	M7	S2	M2
7	M7	S1	R1	M3
8	S1	S2	R2	M4
9	S2	R1	PS1	M5
10	R1	R2	PS2	M6
11	R2	PS1	M1	M7
12	PS1	PS2	M2	S1
13	PS2	M1	M3	S2

Note: M denotes a mathematics item cluster, R a reading item cluster, S a science item cluster and PS a problem solving item cluster. To enable linking between booklets, each of these item clusters appears in four of the test booklets in each of the four possible positions (*i.e.* in Cluster 1, Cluster 2, Cluster 3 or Cluster 4).

TIMSS 2003 booklet design

Student		Part I			Part II	
Booklet	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
1	M1	M2	S6	S7	M5	M7
2	M2	M3	S5	S8	M6	M8
3	M3	M4	S4	S9	M13	M11
4	M4	M5	S3	S10	M14	M12
5	M5	M6	S2	S11	M9	M13
6	M6	M1	S1	S12	M10	M14
7	S1	S2	M6	M7	S5	S7
8	S2	S3	M5	M8	S6	S8
9	S3	S4	M4	M9	S13	S11
10	S4	S5	M3	M10	S14	S12
11	S5	S6	M2	M11	S9	S13
12	S6	S1	M1	M12	S10	S14

Note: M denotes mathematics and S science. To enable linking between booklets, all blocks will appear in at least two of the 12 booklets. Trend items are placed in Part I, and new items are placed in Part II, except for M5 and M6 which appear in both Part I and Part II. Calculators are not allowed for Part I items, but they are allowed for Part II items.

Amount of assessment material

54. Both PISA and TIMSS developed approximately 210 minutes of mathematics assessment material, and these are placed in the test booklets with some duplication across the booklets[9]. Interestingly, in PISA, the 210 minutes of test material were made up of 85 test items (with 94 score points in total), while, in TIMSS, the 210 minutes of test material were made up of 194 items (with 215 score points in

9. The final 2003 tests contain 210 minutes of test material, although many more items were developed and field tested, from which 210 minutes of test material were selected.

total). So, on average, each PISA item should take about 2.5 minutes to complete, while a TIMSS item should take about 1 minute to complete. That is, on average, it is expected that each PISA item would take more than twice the time to complete than a TIMSS item.

55. In PISA, each student took between 30 to 90 minutes of mathematics assessment (with an average of 65 minutes), while, in TIMSS, half of the students took 30 minutes of mathematics assessment, and the other half took 60 minutes of mathematics assessment (with an average of 45 minutes). While, on average, PISA students were given more time to answer mathematics items, the average number of items administered to PISA students was 26 (total of 29 score points), as compared to an average of 42 items (total of 47 score points) administered to each TIMSS student.

56. The reported test reliability for TIMSS is 0.89, and 0.85 for PISA. This is not surprising, as test reliability is closely related to the number of items (or total score points) administered to each student. It appears that PISA developed more extended mathematics tasks in the assessment, but was not able to turn the extended assessment into more score points. In some sense, PISA's attempt to build longer tasks to reflect real-life contexts is at some expense of test reliability.

57. A more detailed comparison of item types in PISA and TIMSS is given in Chapter 3.

Scaling methodology

58. Both PISA and TIMSS use item response theory (IRT) to model student responses to test items. Item response theory is particularly useful for the matrix sampling test design in PISA and TIMSS to "equate" student scores when students took different test booklets, since, in these cases, raw scores on the tests were not directly comparable as different sets of test items were administered to each student. See Box 2.2 for a brief description of IRT.

Box 2.2 Overview of Item Response Modelling

Principles of Item Response Theory (IRT)

Item response theory pertains to the use of mathematical functions to model the probability of success on an item as a function of the characteristics of an item (*e.g.*, difficulty) and characteristics of a person (*e.g.*, ability). Typically, item difficulty parameter and ability parameter are defined on the same measurement scale, so that the relationship between the ability of a person and the difficulty of an item is well defined through a mathematical function. In the case of the one-parameter item response model, the probability of success on an item is given by

$$probability\ of\ success = \frac{\exp(\theta - \delta)}{1 + \exp(\theta - \delta)}$$

where θ is the person's ability on the measurement scale, and δ is the item difficulty on the same measurement scale. When a person's ability is equal to the item difficulty, the probability of success is 0.5. Consequently, if a person is administered a number of test items with known item difficulty parameters, one can make an estimate of the person's ability based on the observed patterns of successes and failures on the items. The advantage of this approach is that the estimate of a person's ability is invariant (within measurement error) of the particular set of items administered. This property of the one-parameter item response model is particularly useful when there is a rotated booklet test design where students take different sets of the items such as the booklets in PISA and TIMSS.

In addition, since person ability and item difficulty are defined on the same measurment scale, one can make statements about a student (located at a point on the scale) with respect to his/her likelihood of being successful on various items which are also located on the scale according to the item difficulty parameter. In this way, one can provide descriptors of skills and knowledge to illustrate typical capcacities of students located at various poinsts on the measurement scale.

There are also many different item response models, where different mathematical functions are used to describe the probability of success. In particular, the two-parameter and three-parameter item response models are also commonly used. The two-parameter item response model has an additional parameter describing the discrimination power of each item. For example, opend-ended items typically provide more discrimination power than multiple-choice items in separating students on the measurement scale (Routitsky & Turner, 2003). The three-parameter model has an additional "guessing" parameter for multiple-choice items.

In general, the choice of a particular item response model is often influenced by different schools of thought. All item response models have theoretical and practical advantages and drawbacks. A detailed discussion on the differences between item response models is beyond the scope of this report. Some references on item response theory include Embretson and Reise, 2000, and van der Linden and Hambleton, 1997.

59. There is, however, a difference in the item response models used in PISA and TIMSS. In PISA, the one-parameter item response model was used (OECD, 2005), while in TIMSS, the three-parameter model was used (IEA, 2004). There appears to be no clearly documented findings comparing the one-parameter and three-parameter item response models. In TIMSS 1995, student responses were modelled using both the one-parameter model and the three-parameter model, but IEA did not publish any direct comparisons of the two scaling methods

60. From a theoretical point of view, the three-parameter item response model takes into account the discrimination power of individual items, that is, to what extent the item can separate poor ability students from high ability students, as well as guessing factors for multiple-choice items. Consequently, if test items are greatly different in terms of discrimination power, it is possible that the two scaling methods could produce a different "spread" of the student ability distributions. The ranking of the countries is likely to be unchanged (up to the accuracy of the proficiency estimates), provided that the items do not exhibit large

differential item functioning (DIF) across countries. While DIF was checked after the field trial in PISA, this is an issue that still remains to be investigated.

61. Both PISA and TIMSS use plausible values methodology (Mislevy, 1991; Mislevy, *et al*, 1992) for estimating student achievement distributions, as well as replication methods for computing standard errors of estimates. TIMSS uses the jacknife replication method, and PISA uses the balanced repeated replication method. Both methods use the same approach in estimating standard errors for complex samples. For more detail, consult the technical reports of PISA and TIMSS. See Box 2.3 for a brief description of the Plausible Values methodology.

Box 2.3 Plausible Values

There are at least two different approaches to estimate student achievement distributions. The first is an indirect approach where an estimate of each student's ability is first made based on the student's item responses, and these ability estimates are then aggregated to form population characteristics. The second approach is a direct estimation method where a parametric model for the ability distribution is assumed (for example, as a normal distribution with mean μ and variance σ^2). The parameters of the distribution are then estimated directly using student item responses. The plausible values methodology is a direct estimation method for population characteristics. Plausible values can be regarded as computational tools for building the population ability distribution. This direct estimation method overcomes some of the problems encountered in the indirect approach where inaccuracies in individual student ability estimates cause biases in population estimates. Readers can refer to Wu (2005b) for a more detailed explanation on plausible values.

Field operation procedures

62. Field operation procedures in PISA and TIMSS are very similar, with provisions for translation verification, marker training and marker reliability studies, national centre monitors, school visits, and a whole suite of quality control procedures to ensure the standards of survey operations across all participating countries. It can be said both PISA and TIMSS adopt similar field operation procedures with only minor variations. For example, in PISA, source documents are prepared in both English and French, and a double translation is carried out using English and French source documents separately. In TIMSS, the source language is in English, and an independent double translation is carried out. Both PISA and TIMSS have independent translation verification processes in place, where international translation companies verify the translated instruments. The respective study centres then review and check for the reconciliation of the various translations.

63. Test administration procedures are also very similar between the two studies, with some minor differences such as lengths of testing time. However, policies on calculator use are somewhat different between the two surveys. Since the use of calculators may have an impact on mathematics achievement, the following section examines this issue more closely.

Calculator use

64. PISA and TIMSS have different rulings on calculator use. PISA adopts an open position on the use of calculators, as the following shows:

> National centres decided whether calculators should be provided for their students on the basis of standard national practice. No items in the pool required a calculator, but some items involved solution steps for which the use of a calculator could facilitate computation. In developing the mathematics items, test developers were particularly mindful to ensure that the items were as calculator-neutral as possible. (p.16, OECD, 2005)

65. TIMSS did not allow the use of calculators in 1995 and 1999. However, in 2003, for Grade 8, calculators were allowed for some items. For Part I of the TIMSS test booklets, calculators were not allowed. For Part II, calculators were allowed. As for PISA, *"TIMSS mathematics items were designed so that they could be answered readily without the use of a calculator (p.374, IEA, 2003)"*.

66. The differences between PISA and TIMSS in the policy on calculator use reflect differences in the nature of the two assessments.

> The PISA assessment focuses on problem situations that arise from the real world, and the use of calculators is very much a part of everyday life (whether at work or at home). However, it should be stressed that intensive computation is not a key focus of the PISA test, and there will not be purely computational items that depend solely on the use of the calculator. (p.365, OECD, 2005)

67. In contrast, TIMSS clearly regards computation skills without the use of calculators as one important strand of mathematics, as reflected in the inclusion of mathematics items where calculators were not allowed. For TIMSS 2003 Grade 4 assessment, no calculators were allowed.

68. The impact of calculator use on achievement results is examined in Chapter 4.

Questionnaires

69. Apart from the mathematics assessment instruments, contextual information about students and schools is also collected in both PISA and TIMSS. This information is not only useful in its own right, it also provides analytic power for gaining an in-depth understanding of achievement results. PISA administers a student questionnaire and a school questionnaire, while TIMSS administers four questionnaires: a student questionnaire, a teacher questionnaire, a school questionnaire and a curriculum questionnaire. Since the sampling design of PISA does not involve the selection of intact classes, it is difficult to administer a teacher questionnaire. As students come from different classes in each school, it becomes difficult to identify all the teachers of sampled students and link students to teachers. Consequently, in PISA, limited amount of teacher information is collected through the student and school questionnaires. In contrast, TIMSS collects extensive teacher information including teacher background, school and classroom climate, instructional approaches and implemented curriculum. In addition, TIMSS administers a Curriculum questionnaire seeking information on the process of curriculum development in each country, as well as the mathematics topics included in each country's curriculum.

Box 2.4 How PISA and TIMSS collected information from students in the 2003 surveys

Student questionnaire in TIMSS at Grade 8

Students took 30 minutes to answer 23 questions in total. There were seven main sections:

About you - questions about the students and their family, including date of birth, sex, number of books at home, home possessions (16 items, of which 12 are country specific), language spoken at home, educational level of parents and their own educational expectations.

Mathematics in school – students' attitudes towards learning mathematics, how frequently students are taught or learn in different suggested ways in mathematics lessons (14 different ways).

Science in school – students' attitudes towards learning science, how frequently students are taught or learn in different suggested ways in science lessons (14 different ways).

Computers – where students use computers and how frequently they complete 4 suggested educational tasks.

Your school – students' attitudes towards school and teachers (4 statements) and school climate and sense of belonging (5 statements).

Things you do outside of school – how frequently students spend time on 9 selected activities before or after school (including leisure pursuits and both school and paid work), frequency of extra lessons or tutoring and homework in mathematics and science.

More about you – how many people live at the students' home and whether or not the students and their parents or guardians were born in the country of assessment and if applicable at what age the students arrived in the country of assessment.

PISA student questionnaires

PISA administers a main student questionnaire, as well as providing countries the option of administering additional shorter questionnaires to collect information such as students' use of computers. There were three possible questionnaires for students to complete in 2003. All participating students took about 30 minutes to answer 38 questions in the main student questionnaire. There were six main sections:

About you – Grade at school, type of educational programme, date of birth, sex

You and your family – who lives at home with the students, parents' employment status, occupation (including job title and tasks) and education (level completed and qualifications), country of birth of both student and parents, and, if applicable, at what age the student arrived in the country of assessment; language most often spoken at home; home possessions (16 items, including a computer, of which 3 are country specific); number of 5 selected home possessions; number of books at home.

Your education – pre-primary school attendance, age student started primary school, grade repetition if applicable, level of education student expects to complete, attitudes towards school.

Your school – reasons student attends the school, views on teachers in the school, sense of belonging at school, punctuality, number of hours spent on homework, extra classes or with tutors outside school.

Learning mathematics – views on mathematics learning (8 statements), confidence in mathematics (8 tasks), studying mathematics (10 statements), time spent on learning mathematics, different ways of studying mathematics (14 statements).

Your mathematics classes – number and length in minutes of mathematics classes per week, number of students in the mathematics class, cooperative and competitive ways of learning mathematics (10 statements), mathematics classroom environment, including disciplinary climate and teacher support (11 statements).

In addition, students in 21 countries took five minutes to complete eight questions in an optional questionnaire on Educational career.

Educational career - Absence of at least two months and/or change of school in primary (ISCED 1) or lower secondary (ISCED 2) school; change of study programme in current grade; type of mathematics class; mathematics mark in last school report and whether it was above or below the pass mark; and expected job at the age of 30.

Students in 32 countries took five minutes to complete nine questions in an optional questionnaire on Information and Communication Technology (ICT).

ICT - Availability of computers at home, school and elsewhere and how frequently students use these; the number of years students have been using a computer; how frequently students use computers for specific tasks (12 tasks); students' confidence in completing different tasks on a computer (23 tasks); students' attitudes towards using computers; and how students learned to use computers and the Internet.

70. The extent to which student, teacher and school background information is collected has an impact on the analyses that can be carried out. Box 2.3 presents an overview of the questions that students answered in PISA and TIMSS in 2003. Both PISA and TIMSS examine students' "self-confidence" and "interest and motivation" in mathematics. A comparison of these results is given in Chapter 5. PISA also collects information on students' learning strategies and how learner characteristics influence mathematics performance. TIMSS international report has a brief presentation of students' socio-economic status (SES), while PISA devotes a large section of the report to the analysis of SES in relation to achievement. TIMSS presents a detailed report on the profiles of teachers, including cross-country comparisons of teacher qualifications, gender, age, experience, and professional activities and support for teachers.

71. Both PISA and TIMSS report on school contexts for learning, including student-teacher relationships, school resources, and classroom climate. TIMSS further presents details of the mathematics classroom in terms of resources used in mathematics classes, content taught, and assessment methods. The main difference between PISA and TIMSS in relation to contextual information is that classroom level information is collected from teachers directly in TIMSS, while, in PISA, information is aggregated within each school from students' responses to questions about the classroom environment.

Summary

72. The following is a summary of similarities and differences between PISA and TIMSS Grade 8 discussed in this Chapter.

Aspects of Survey	Specific Point	Similarities	Differences
Sampling	Population versus sample	Both surveys are sample-based	
	Population definition		PISA is age-based. TIMSS is grade-based
	Age distribution		Typically, there is a two-year age span in each country's sample of students in TIMSS, and one-year age span in PISA.
	Grade distribution		Typically, there are two grades involved in each country in the PISA sample, and one grade in TIMSS.
	Sampling method	Both surveys use a two-stage sampling method, where schools are first selected using probability proportional to size method.	In TIMSS, the second stage of sampling selects intact classes. In PISA, the second stage of sampling selects students at random within each school.

Aspects of Survey	Specific Point	Similarities	Differences
	Sample size	Both PISA and TIMSS require a minimum of 150 schools to be selected. PISA recommends that the total number of sampled students is at least 5250, while TIMSS requires a minimum of 4000 students to be selected in each country.	
Test Characteristics	Test length		73. The testing time is two hours in PISA and 90 minutes in TIMSS.
	Test design	Both surveys use rotated test booklet design. PISA uses 13 booklets. TIMSS uses 12 booklets.	
	Amount of assessment material	Both PISA and TIMSS developed approximately 210 minutes of mathematics assessment material	There are 94 total score points in PISA, and 215 total score points in TIMSS.
Scaling Methodologies	Item response model	Both surveys use item response modelling and plausible values methodologies for estimating student ability distributions	PISA uses the one-parameter item response model. TIMSS used the three-parameter item response model.
Field Operations	Translation	Both PISA and TIMSS have translation verification process in place.	In PISA, source documents are prepared in both English and French, and a double translation is carried out using English and French source documents separately. In TIMSS, the source language is in English, and an independent double translation is carried out.
	Quality monitor	Both PISA and TIMSS have similar procedures for marker training and marker reliability studies, national centre monitors and school visits.	

Aspects of Survey	Specific Point	Similarities	Differences
Test administration	Calculator use		In PISA, calculators are allowed. In TIMSS, calculators are allowed for only Part II of the test. More students have access to calculators in PISA than in TIMSS.

CHAPTER 3 - COMPARISON OF PISA AND TIMSS MATHEMATICS FRAMEWORKS AND ITEM FEATURES

Approaches to the development of the mathematics assessment frameworks

74. PISA and TIMSS adopted different approaches to the development of the assessment frameworks. Each survey developed the assessment framework to meet their objectives which are somewhat different. In PISA, the aim is to assess the extent to which education systems have prepared 15-year-olds "to play constructive roles as citizens in society" (p.24, OECD, 2003), so the assessment focuses on "what are the skills citizens require to play constructive roles in society?". In TIMSS, the assessment is to improve teaching and learning of mathematics, so the assessment provides information about student achievement levels in relation to what students have learned in schools. This difference in the orientation of the purposes of the two assessments led to the different approaches to the development of the frameworks. A review of each framework is given below, followed by a comparison between the two frameworks.

TIMSS mathematics assessment framework

75. The overall design of TIMSS evolves around the TIMSS Curriculum Model where three levels of curriculum, the intended curriculum, the implemented curriculum and the attained curriculum, form the major organising principle of the TIMSS study (p.3. IEA, 2003b). Questionnaires designed for students, teachers and school principals aim to capture the first two levels of curriculum structure, namely, the intended curriculum and the implemented curriculum, while the assessment attempts to capture the attained curriculum. TIMSS stresses that the usefulness of the TIMSS results to policy makers "depends on achievement measures being based, as closely as possible, on what students in their systems have actually been taught" (p.5. IEA, 2003b).

76. To ensure that the test contents are aligned with what students were taught in the participating countries, a survey was conducted to collect information on the curricula in participating countries. Mathematics topics that were regarded as important in a significant number of countries were included in the framework. However, TIMSS stresses that "the frameworks do not consist solely of content and behaviours included in the curricula of all participating countries" (p.5., IEA, 2003b). The following six factors underlie the principles of the inclusion of mathematics content domains in the assessment framework (p.5, IEA, 2003b):

- Inclusion of the content in the curricula of a significant number of participating countries;
- Alignment of the content domains with the reporting categories of TIMSS 1995 and TIMSS 1999;
- The likely importance of the content to future developments in mathematics and science education;
- Appropriateness for the populations of students being assessed;
- Suitability for being assessed in a large-scale international study;
- Contribution to overall test balance and coverage of content and cognitive domains.

77. Of the six factors listed above, the third one does not quite fit in with the principle that the achievement measures reflect what students have actually been taught, since there is an implication that, if a topic is deemed important for future developments in mathematics, then it may be included whether or not students have been taught the topic. From this point of view, the TIMSS framework is not completely driven by national curricula. The framework also seeks to set some directions for future directions in mathematics education.

78. There are two organising dimensions underlying TIMSS 2003 framework: Mathematics content domains and mathematics cognitive domains. These are discussed separately below.

TIMSS mathematics content domains

79. Box 3.1 lists the five mathematics content domains in the TIMSS framework. *Number*, *algebra*, *measurement*, *geometry* and *data* are familiar labels to mathematics educators, as they mirror closely the content domains that are often found in the mathematics curricula of most countries. For more detail, refer to the TIMSS 2003 Frameworks document (IEA, 2003b).

Box 3.1 The TIMSS mathematics content domains

There are five main content domains in the TIMSS 2003 mathematics assessment:

Number

Whole numbers
Fractions and decimals
Integers
Ratio, proportion and percent

Algebra

Patterns
Algebraic expressions
Equations and formulas
Relationships

Measurement

Attributes and units
Tools, techniques and formula

Geometry

Lines and angles
Two- and three-dimensional shapes
Congruence and similarity
Locations and spatial relationships
Symmetry and transformations

Data

Data collection and organisation
Data representation
Data interpretation
Uncertainty and probability

80. The TIMSS framework document includes a list of topics covered by each content domain (see Box 3.1), and, within each topic, a set of assessment outcomes to illustrate the specific tasks that students will typically be assessed on.

81. For example, Box 3.1 lists four topics for the content area, *number*: whole numbers, fractions and decimals, integers, and ratio, proportion, and percent. Within the topic area of ratio, proportion, and percent, the assessment outcomes (topic bullets) are:

- Identify and find equivalent ratios.
- Divide a quantity in a given ratio.
- Convert percents to fractions or decimals, and vice versa.
- Solve problems involving percents.
- Solve problems involving proportions.

82. The TIMSS framework appears very comprehensive in its lists of the main topics and assessment outcomes within each content area. A check of the actual items in the TIMSS 2003 tests showed that, out of a total of 19 topics, only one topic (Data collection and organisation in the *data* content domain) was not assessed in the TIMSS 2003 tests. Out of a total of 87 assessment outcomes (topic bullets), 67 were covered by items in the actual tests[10]. However, some topic bullets were assessed by numerous items, while others were assessed by only one item. The proportions of items in different content domains are given in Table 3.1.

Table 3.1 Number and proportions of items in TIMSS by content domain

	No. of Items in TIMSS 2003 tests	Proportion of items	TIMSS Target proportion of items	International average of % of time taught in schools
Number	57	30%	30%	21 %
Algebra	47	24%	25%	27%
Measurement	31	16%	15%	10%
Geometry	31	16%	15%	26%
Data	28	14%	15%	10%
Total	**194**	**100%**	**100%**	**94%**[1]

1. The total does not equal 100%, as 6% of the content domains reported by teachers are not covered by the five areas listed in the table.

83. The last column in Table 3.1 shows the international average of the percentage of time in mathematics class devoted to each TIMSS content area during the school year, as reported by teachers[11]. There is some discrepancy between the proportions of items in TIMSS tests and the average percentages of time the content domains are taught across all TIMSS participating countries. The TIMSS mathematics framework does not use content coverage as the sole criterion for determining the relative weights of the

10 . We obtained slightly different figures depending on whether we used the item list provided by the Australian TIMSS National Research Coordinator, or the item list in the released data set.

11 . Figures are obtained from Exhibit 7.4 of the TIMSS 2003 International Mathematics Report (IEA, 2003).

content domains. Rather, the TIMSS framework "represents a consensus among the countries participating in TIMSS 2003 about the mathematics students at these grades should be expected to have learned." (IEA, 2003, p.180)

TIMSS mathematics cognitive domains

84. TIMSS mathematics cognitive domains relate to the types of cognitive skills required in doing mathematics, generic across all mathematics content domains. To achieve a balanced test, it is desirable to ensure that each type of skills and abilities is covered by a sufficient number of items. There are four cognitive domains in TIMSS 2003:

- Knowing facts and procedures
- Using concepts
- Solving routine problems
- Reasoning

85. These four cognitive domains are listed in order of the complexity of the tasks, from straightforward problems to complex tasks. However, the TIMSS framework stresses that cognitive complexity should not be confused with item difficulty, in that there is a range of item difficulties associated with each cognitive domain (IEA, 2003b, p.25. That is, within each cognitive domain, there are easy items as well as difficult items. As these labels of cognitive domains are not necessarily familiar to the reader, a brief description of each cognitive domain is given below.

Knowing facts and procedures

86. This cognitive domain covers basic language of mathematics and essential mathematical facts and properties, as well as the use of mathematics for solving routine problems typically encountered in everyday life. There are four categories of skills covered by this cognitive domain:

- Recall – *e.g.*, knowing number facts, mathematical conventions/notations.
- Recognise/identify – *e.g.,* recognising different representations of the number system.
- Compute – *e.g.,* carrying out arithmetic computation, expanding algebraic expressions.
- Use tools – *e.g.,* reading scales, using straightedge and compass.

Using concepts

87. This cognitive domain is about the ability to make connections of knowledge, judge the validity of mathematical statements and create mathematical representations. There are five categories of skills under this cognitive domain:

- Know – *e.g.,* knowing concepts such as inclusion and exclusion, generality, mathematical relationships.
- Classify – *e.g.,* grouping objects, shapes, numbers according to common properties.
- Represent – *e.g.,* presenting information using tables, diagrams, graphs; moving between equivalent representations of mathematical relationships.
- Formulate – *e.g.,* modelling problems or situations with equations or expressions.

- Distinguish – *e.g.,* identifying valid and invalid inferences from questions and answers.

Solving routine problems

88. This cognitive domain relates to problem solving, where the problems are routine in that they are typically encountered as classroom exercises or in textbooks. There are five categories of skills identified under this cognitive domain:

- Select – *e.g.,* choosing an appropriate algorithm or strategy to solve a problem
- Model – *e.g.,* generating an appropriate model using equations or diagrams.
- Interpret – *e.g.,* understanding a given model presented as equations or diagrams.
- Apply – *e.g.,* using knowledge of facts, procedures, and concepts to solve routine problems.
- Verify/Check – *e.g.,* Checking and evaluating the correctness and reasonableness of the solution

Reasoning

89. This cognitive domain is about solving non-routine problems using logical, systematic thinking and various forms of reasoning. Eight categories of skills have been identified in relation to this cognitive dimension:

- Hypothesize/Conjecture/Predict – *e.g.* discussing ideas, specifying an outcome resulting from an unperformed operation
- Analyze – *e.g.* making valid inferences from given information, decomposing geometric figures
- Evaluate – *e.g.* critically evaluating mathematical ideas, methods, etc.
- Generalize – *e.g.* restating results in more widely applicable terms
- Connect – *e.g.* linking related mathematical ideas or objects
- Synthesize/Integrate – *e.g.* combining results to solve a problem
- Solve non-routine problems – *e.g.* applying mathematical procedures in unfamiliar contexts
- Justify/Prove – *e.g.* providing evidence for the validity of a statement using mathematical results

90. The proportions of items classified by the TIMSS cognitive domains are given in Table 3.2.

Table 3.2 Proportions of items in TIMSS by cognitive domains

	Proportion of items in TIMSS tests	Target proportion of items
Knowing facts and procedures	23%	15%
Using concepts	19%	20%
Solving routine problems	36%	40%
Reasoning	22%	25%
Total	**100%**	**100%**

PISA mathematics assessment framework

91. The PISA mathematics framework begins with a formal definition of mathematical literacy for OECD/PISA:

> Mathematical literacy is an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgements and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen. (p. 24, OECD, 2003)

92. This definition, when stated alone, does not show how it differs from the mathematics construct of TIMSS. In fact, the TIMSS framework states the following:

> Prime reasons for inclusion of mathematics (in school curricula) are the increasing awareness that effectiveness as citizens and success in the workplace are greatly enhanced by knowing and, more important, being able to use mathematics. The number of vocations that demand a high level of proficiency in the use of mathematics, or mathematical modes of thinking, has burgeoned with the advance of technology, and with modern management methods. (p.30, IEA 2003b)

93. What makes the PISA mathematics framework different from the TIMSS framework, and different from typical mathematics curricula in most countries, is the fact that PISA does not make the assumption that school mathematics will necessarily prepare students to be mathematically literate in their future lives as effective citizens. This is evident in the following:

> Rather than being limited to the curriculum content students have learned, the assessments focus on determining if students can use what they have learned in the situations they are likely to encounter in their daily lives. (p.24, OECD, 2003)

94. That is, PISA sets out to establish the mathematical knowledge and skills required to be mathematically literate citizens, and assesses students on these. This is a more direct way to obtain measures of whether students can meet future challenges in life, rather than via a proxy that school achievement in mathematics is an indicator of students' capacity to use mathematics to solve everyday problems. This orientation may have come about in recent years when mathematics educators observed that many students regarded school mathematics as an academic discipline divorced from real life (*e.g.*, Bonotto, 2003; Verschaffel, Greer & de Corte, 2000). Further, the theory of Realistic Mathematics Education (RME) developed in the Netherlands (de Lange, 1996; Gravemeijer, 1999) over the past 30 years has gathered support from around the world (de Lange, 1996; Romberg & de Lange, 1998). Two principles underlie RME: (1) Mathematics must be connected to the real-world; (2) Mathematics should be seen as a human activity.

95. PISA has not gone so far as claiming that schools are not preparing students adequately to be mathematically literate citizens. But the approach PISA has adopted does not make the assumption that schools do prepare students well. One would hope that the different approaches to organising the mathematics assessment frameworks in PISA and TIMSS would not lead to different results, since one of the main aims of schooling must be to prepare students for their wellbeing in their future lives. If the two surveys yield different results that cannot be explained by methodological differences, a close examination would be called for to understand the differences.

96. The PISA approach to defining mathematical literacy stems from the definition of "literacy" in James Gee's "Preamble to a literacy program" (1998) where literacy refers to the use of language. Each human language has words and rules, but to use a language effectively, one needs to know how to combine words and rules in complex ways to convey a vast array of ideas. Analogous to language, mathematics also consists of building blocks such as symbols, terms and rules. But to use mathematics effectively, one needs to know how to combine the building blocks of mathematics to solve specific problems. That is, mathematics literacy is much more than just knowing mathematical symbols and

rules. Mathematics literacy is about how people use their mathematical knowledge to actually solve real-world problems.

97. PISA identifies three dimensions for the organisation of the mathematics framework. These dimensions are: (1) situation or context; (2) mathematical content; and (3) mathematical processes. These three dimensions are described below.

Situation/context dimension

98. The following paragraph from the PISA framework describes the situation/context dimension:

> The situation is the part of the student's world in which the tasks are placed. It is located at a certain distance from the students. For OECD/PISA, the closest situation is the students' personal life; next is school life, work life and leisure, followed by the local community and society as encountered in daily life. Furthest away are scientific situations. Four situation-types will be defined and used for problems to be solved: personal, educational/occupational, public, and scientific. (p. 32, OECD, 2003).

99. The PISA framework specifies that, as far as possible, the target proportions of each situation/context type should be about equal. The actual proportions of items by situation/context in the PISA assessment are given in Table 3.3.

Table 3.3 Number and proportions of mathematics items in PISA 2003 by situation/context dimension

	Number of items	Proportion of items
Personal	18	21%
Educational/occupational	20	24%
Public	29	34%
Scientific	18	21%
Total	**85**	**100%**

100. The PISA framework also discusses situation/context from the point of view of the distance between the problem and the mathematics involved. Tasks involving only mathematical objects without any reference to matters outside the mathematical world are termed "intra-mathematical", while tasks involving real-world objects are termed "extra-mathematical". PISA places an emphasis on extra-mathematical tasks. Of the 85 items, only one was classified as intra-mathematical (and this one was also classified as "scientific" in Table 3.3), while all other items were extra-mathematical. That is, while PISA does not preclude intra-mathematical tasks in the framework, the items in the actual assessment are essentially all extra-mathematical.

PISA mathematical content dimension – The overarching ideas

101. PISA adopts a phenomenological organisation for mathematical content as described below:

> Since the goal of OECD/PISA is to assess students' capacity to solve real problems, our strategy has been to define the range of content that will be assessed using a phenomenological approach to describing the mathematical concepts, structures or ideas. This means describing content in relation to the phenomena and the kinds of problems for which it was created. This approach ensures a focus in the assessment that is consistent with the domain definition, yet covers a range of content that includes what is typically found in other mathematics assessments and in national mathematics curricula. (p.34, OECD, 2003).

102. In other words, PISA considers the world around us, and categorises the tasks that are typically encountered in everyday life, and uses these categories as the basis for organising the mathematics content. This approach possibly explains why there is only one intra-mathematical item in the PISA test, since intra-mathematical tasks are not often encountered in most people's everyday life, outside the school environment.

103. Interestingly, PISA appears to suggest that the phenomenological approach is more inclusive than school curriculum, as in the last sentence of the above quote, "This approach … covers a range of content that *includes* what is typically found … in national mathematics curricula." But from the point of view of intra-mathematical and extra-mathematical tasks, and that PISA focuses on individual's "everyday life" instead of focusing on mathematics as in "how mathematics is used in the world", it would appear that PISA's approach results in a subset of the tasks found in national mathematics curricula. This point will be further discussed later in this report.

104. On the other hand, PISA's approach could be viewed as more inclusive from the point of view that the tasks often involved skills from multiple (traditional) content domains, and no isolated knowledge or skill is tested without checking whether these skills can be applied to real-life situations.

105. It should be noted that PISA makes a distinction between approaches to assessment and teaching, as the following paragraph shows:

> Mathematical concepts, structures and ideas have been invented as tools to organise the phenomena of the natural, social and mental world. In schools, the mathematics curriculum has been logically organised around content strands (*e.g.*, arithmetic, algebra, geometry) and their detailed topics that reflect historically well-established branches of mathematical thinking, and that facilitate the development of a structured teaching syllabus. (p. 34, OECD, 2003)

106. It is important to recognise that, while PISA organises the mathematic content differently from typical school mathematics curriculum, PISA does not suggest that the organisation based on phenomenological approach is necessarily appropriate for organising a structured teaching syllabus. Clearly, students cannot be taught tasks involving skills from multiple content domains without having been taught basic building blocks of mathematics knowledge and procedures in each content domain. This distinction between assessment and teaching is an important one, in that the comparison between PISA and TIMSS is focused on assessment, and not on teaching, although there is a strong relationship between the two.

107. PISA's phenomenological approach to organising mathematics content identifies four areas, called *overarching ideas*. The four overarching ideas are *quantity, space and shape, change and relationships, uncertainty*. The PISA mathematics framework provides the following descriptions for each of the four overarching ideas (OECD, 2003, pp. 36-37).

Quantity

108. This overarching idea focuses on the need for quantification in order to organise the world. Important aspects include an understanding of relative size, the recognition of numerical patterns, and the use of numbers to represent quantities and quantifiable attributes of real-world objects (counts and measures). Furthermore, *quantity* deals with the processing and understanding of numbers that are represented to us in various ways.

109. An important aspect of dealing with *quantity* is quantitative reasoning, Essential components of quantitative reasoning are number sense, representing numbers in various ways, understanding the meaning of operations, having a feel for the magnitude of numbers, mathematically elegant computations, mental arithmetic and estimating.

Space and shape

110. Patterns are encountered everywhere: in spoken words, music, video, traffic, building constructions and art. Shapes can be regarded as patterns: houses, office buildings, bridges, starfish, snowflakes, town plans, cloverleaves, crystals and shadows. Geometric patterns can serve as relatively simple models of many kinds of phenomena, and their study is possible and desirable at all levels (Grünbaum, 1985).

111. The study of shape and constructions requires looking for similarities and differences when analysing the components of form and recognising shapes in different representations and different dimensions. The study of shapes is closely connected to the concept of "grasping space". This means learning to know, explore and conquer, in order to live, breathe and move with more understanding in the space in which we live (Freudenthal, 1973).

112. To achieve this requires understanding the properties of objects and their relative positions. We must be aware of how we see things and why we see them as we do. We must learn to navigate through space and through constructions and shapes. This means understanding the relationship between shapes and images or visual representations, such as that between a real city and photographs and maps of the same city. It includes also understanding how three-dimensional objects can be represented in two dimensions, how shadows are formed and must be interpreted, what perspective is and how it functions.

Change and relationships

113. Every natural phenomenon is a manifestation of change, and the world around us displays a multitude of temporary and permanent relationships among phenomena. Examples are organisms changing as they grow, the cycle of seasons, the ebb and flow of tides, cycles of unemployment, weather changes and stock exchange indices. Some of these change processes involve and can be described or modelled by straightforward mathematical functions: linear, exponential, periodic or logistic, either discrete or continuous. But many relationships fall into different categories, and data analysis is often essential to determine the kind of relationship that is present. Mathematical relationships often take the shape of equations or inequalities, but relations of a more general nature (*e.g.*, equivalence, divisibility, inclusion, to mention but a few) may appear as well.

114. Functional thinking – that is, thinking in terms of and about relationships – is one of the most fundamental disciplinary aims of the teaching of mathematics (MAA, 1923). Relationships may be given a variety of different representations, including symbolic, algebraic, graphical, tabular and geometrical. Different representations may serve different purposes and have different properties. Hence translation between representations often is of key importance in dealing with situations and tasks.

Uncertainty

115. The present "information society" offers an abundance of information, often presented as accurate, scientific and with a degree of certainty. However, in daily life we are confronted with uncertain election results, collapsing bridges, stock market crashes, unreliable weather forecasts, poor predictions for population growth, economic models that don't align, and many other demonstrations of the uncertainty of our world.

116. Uncertainty is intended to suggest two related topics: data and chance. These phenomena are respectively the subject of mathematical study in statistics and probability. Relatively recent recommendations concerning school curricula are unanimous in suggesting that statistics and probability should occupy a much more prominent place than has been the case in the past (Committee of Inquiry into the Teaching of Mathematics in Schools, 1982; LOGSE, 1990; MSEB, 1990; NCTM, 1989; NCTM, 2000)..

117. Specific mathematical concepts and activities that are important in this area are collecting data, data analysis and display/visualisation, probability and inference.

118. Table 3.4 shows the number and proportion of PISA items classified according to the *overarching ideas.*

Table 3.4 Number and proportions of mathematics items in PISA 2003 by *overarching ideas*

	Number of items	Proportion of items
Quantity	23	27.0%
Space and shape	20	23.5%
Change and relationships	22	26.0%
Uncertainty	20	23.5%
Total	**85**	**100.0%**

Mathematical processes dimension

119. The PISA mathematics framework deems the mathematical processes dimension to be the most important one, and devotes lengthy discussions to it. As an introduction to describing the mathematical processes dimension, the PISA framework begins with a description of the process of mathematisation which characterises the way mathematics problems are solved in the real world (OECD, 2003, p.27).

120. There are five steps that characterise the process of mathematisation (OECD, 2003, p.38):

1. Starting with a problem situated in reality
2. Organising it according to mathematical concepts
3. Gradually trimming away the reality through processes such as making assumptions about which features of the problem are important, generalising and formalising (which promote the mathematical features of the situation and transform the real problem into a mathematical problem that faithfully represents the situation
4. Solving the mathematical problem
5. Making sense of the mathematical solution in terms of the real situation.

121. The cycle of mathematisation involves an iterative process of moving between a real-world problem to a mathematical problem, and then moving from a mathematical problem to a mathematical

solution and then to a real solution. The process ends with making reflections of the solutions in terms of the real-world problem.

The Competencies

122. To be able to successfully carry out the mathematisation process, an individual will need to draw upon a number of competencies. PISA mathematics framework identifies eight competencies in relation to the mathematisation process:

- Thinking and reasoning
- Argumentation
- Communication
- Modelling
- Problem posing and solving
- Representation
- Using symbolic, formal and technical language and operations
- Use of aids and tools

123. Note that each of the above competencies can be described at different levels. The mathematisation process required to solve a particular problem may draw upon different competencies at different levels.

124. While an explication of these competencies is useful in teaching and learning, it is difficult to assess these competencies individually, since problem-solving tasks typically involve a combination of the competencies, and, in a large-scale assessment, the behaviours of students in relation to each competency would be difficult to observe. Consequently, PISA summarises the competencies into three broad clusters: the reproduction cluster, the connections cluster, and the reflection cluster. Each cluster involves all eight competencies, but at different levels.

125. The PISA framework provides the following definitions for the three competency clusters.

The reproduction cluster

> The competencies in this cluster essentially involve reproduction of practised knowledge. They include those most commonly used in standardised assessments and classroom tests. These competencies are knowledge of facts and of common problem representations, recognition of equivalents, recollection of familiar mathematical objects and properties, performance of routine procedures, application of standard algorithms and technical skills, manipulation of expressions containing symbols and formulae in standard form, and carrying out computations. (p.42, OECD, 2003)

The connections cluster

> The *connections* cluster competencies build on the *reproduction* cluster competencies in taking problem solving to situations that are not simply routine, but still involved familiar, or quasi-familiar, settings. (p.43, OECD, 2003)

The reflection cluster

> The competencies in this cluster include an element of reflectiveness on the part of the student about the processes needed or used to solve a problem. They relate to students' abilities to plan solution strategies and implement them in problem settings that contain more elements and may be more "original" (or unfamiliar) than those in the *connections* cluster. (p.46, OECD, 2003)

Classifying items according to competency clusters

126. Each PISA item was classified into one of the three competency clusters. The classification process involved an examination of the levels of the eight competencies required to answer an item. An item is assigned to a competency cluster according to the highest level of the competencies required. The following table shows the number of PISA items in each of the competency clusters.

Table 3.5 Number and proportions of mathematics items in PISA 2003 by competency clusters

	Number of items	Proportion of items
Reproduction	26	31%
Connection	40	47%
Reflection	19	22%
Total	**85**	**100%**

A Comparison of PISA and TIMSS Mathematics Frameworks

127. TIMSS mathematics framework identifies two dimensions to ensure coverage of mathematics assessment tasks: content domains and cognitive domains. PISA mathematics framework identifies three dimensions for coverage of mathematics assessment tasks: situation/context dimension, content dimension and processes dimension. Of these three dimensions, PISA's content dimension can be related to TIMSS' content domains, while PISA's processes dimension can be related to TIMSS' cognitive domains.

128. TIMSS does not explicitly state a situation/context dimension as PISA does. This is not surprising, since PISA stresses on assessing students' capacity to solve problems encountered in life, PISA items are typically embedded within some situation/context. In contrast, to test knowledge and skills based on curriculum topics, many TIMSS items do not involve matters outside the mathematical world ("intra-mathematical"). This is an important distinction between the assessments of PISA and TIMSS.

A comparison of PISA's content dimension with TIMSS' content dimension

129. TIMSS' content dimension identifies five content domains: *number*, *algebra*, *measurement*, *geometry* and *data*. These content domains are familiar to most mathematics educators as many school mathematics curricula and textbooks are organised around these content domains. In contrast, PISA's content dimension consists of four *overarching ideas*: *quantity*, *space and shape*, *change and relationships*, and *uncertainty*. This classification is less familiar to mathematics educators. To facilitate comparisons between PISA *overarching ideas* and TIMSS content domains, PISA items were classified according to TIMSS content domains, as shown in Table 3.6. Discussions about the relationships between each PISA *overarching idea* and the TIMSS content domains are given following Table 3.6.

Number of PISA items by TIMSS content domains

Table 3.6 Tally of PISA items classified by PISA overarching ideas and TIMSS content domains

		PISA overarching ideas				
		Quantity	**Space and shape**	**Change and relationships**	**Uncertainty**	**Total**
TIMSS content domains	**Number**	23	1	3	5	**32**
	Algebra			7		**7**
	Measurement		6	2		**8**
	Geometry		12			**12**
	Data		1	10	15	**26**
	Total	**23**	**20**	**22**	**20**	**85**

Quantity

130. This *overarching idea* is not dissimilar to the *number* strand in TIMSS. It can be seen from Table 3.6 that all 23 *quantity* items were classified as TIMSS *number* content domain. However, there are quite a few non *quantity* items also classified as TIMSS *number* strand. That is, it appears that the PISA *quantity* domain is a subset of TIMSS *number* domain. Box 3.2 shows a PISA *quantity* item that has been classified as *number* against TIMSS content domains.

Box 3.2 An example PISA *quantity* item classified as *number* against TIMSS content domains

EXCHANGE RATE

Mei-Ling from Singapore was preparing to go to South Africa for 3 months as an exchange student. She needed to change some Singapore dollars (SGD) into South African Rand (ZAR).

Question 1: EXCHANGE RATE

M413Q01 - 0 1 9

Mei-Ling found out that the exchange rate between Singapore dollars and South African Rand was:

1 SGD = 4.2 ZAR

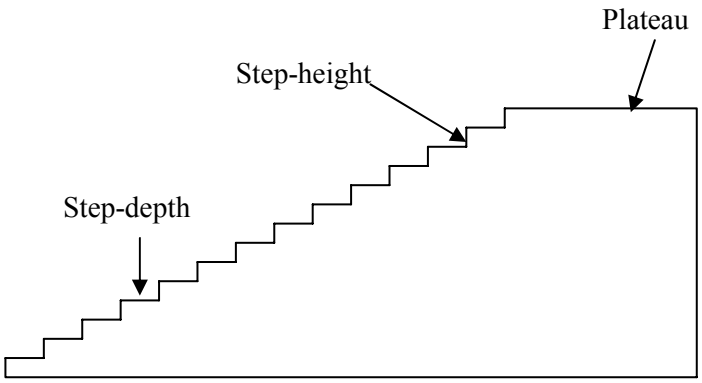
Mei-Ling changed 3000 Singapore dollars into South African Rand at this exchange rate.

How much money in South African Rand did Mei-Ling get?

131. Box 3.3 shows a PISA item that has been classified as TIMSS *number* content domain, but as PISA *space and shape* overarching idea. For this item, while the underlying operation to be carried out is division (which is part of the *number* content domain), but the mathematisation process relates to dealing with space and shape to identify the relevant information before the division operation could be carried out.

Box 3.3 An example PISA *quantity* item classified as *number* against TIMSS content domains

STAIRCASE



The diagram below illustrates a staircase with 14 steps and a total height of 252 cm:

Question 1: STAIRCASE

M547Q01

What is the height of each of the 14 steps?

Height = ... cm.

Space and shape

132. Judging from the PISA items that have been classified as *space and shape*, it appears that this overarching idea covers some topics of *geometry* and *measurement* as defined by the TIMSS framework, for example, two- and three-dimensional shapes and estimates of length, circumference, area and volume. However, checking through the list of TIMSS topics under *measurement* and *geometry*, it appears that many topics are not cover by PISA *space and shape* domain, such as those listed under *lines and angles*, or under *congruence and similarity.* In TIMSS, *measurement* and *geometry* cover a significant number of topics of formal definitions and operations involving lines, angles, polygons, Euclidean and Coordinate Geometry. Since these are often intra-mathematical, PISA does not seem to cover these kinds of knowledge and skills. A check in Table 3.6 shows that most of the PISA *space and shape* items are classified as *geometry* or *measurement* under TIMSS classification scheme. But not all TIMSS *geometry* and *measurement* items are included under PISA *space and shape* overarching idea.

133. Box 3.4 shows a PISA item classified as *space and shape* in PISA, and as *measurement* against TIMSS content domains.

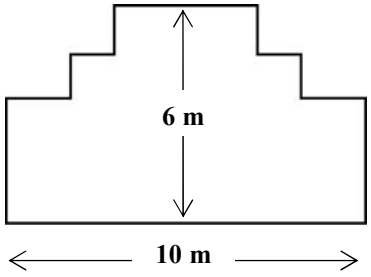
Box 3.4 An example PISA *space and shape* item classified as *measurement* against TIMSS content domains

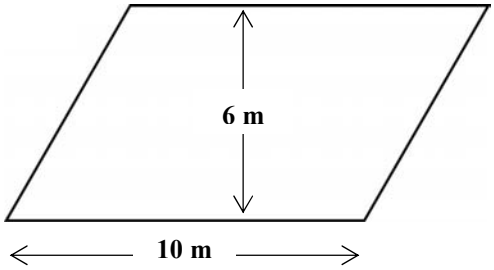
CARPENTER

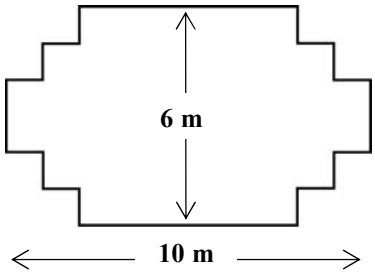
Question 1: CARPENTER

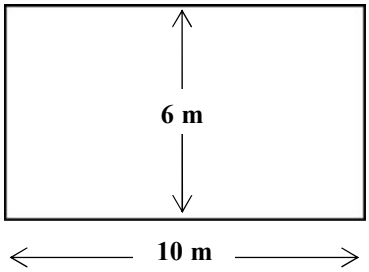
M266Q01

A carpenter has 32 metres of timber and wants to make a border around a garden bed. He is considering the following designs for the garden bed.









Circle either Yes or No for each design to indicate whether the garden bed can be made with 32 metres of timber.

Garden bed design	Using this design, can the garden bed be made with 32 metres of timber?
Design A	Yes / No
Design B	Yes / No
Design C	Yes / No
Design D	Yes / No

Change and relationships

134. From the definitions for this *overarching idea*, it seems reasonable to map the *overarching idea* of *change and relationships* to the traditional curriculum strand algebra. Indeed, the four PISA items classified as TIMSS content domain *algebra* are *change and relationships* items in PISA (see Table 3.6).

However, a large number of items classified as *change and relationships* in PISA are classified by other traditional strands, with the most number in the data strand. This may not be surprising as the descriptions linking *change and relationships* to natural phenomenon include statistical data such as for unemployment or the stock exchange. Looking down the column of *change and relationships* in Table 3.6, the items appear as *number*, *algebra*, *measurement*, and *data* items by TIMSS content domains. This shows that the overarching idea, *change and relationships*, plays a part in most of the traditional mathematics strands. From this point of view, *change and relationships* is probably the least well matched *overarching idea* among the four to traditional curriculum strands.

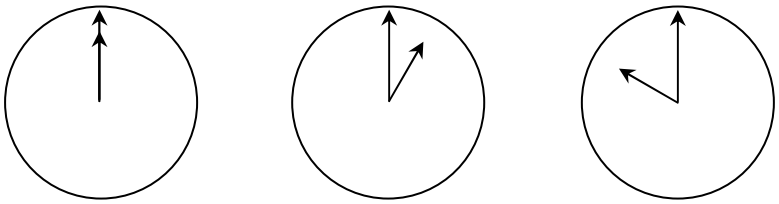
135. Box 3.5 shows two PISA *change and relationships* items classified as TIMSS *measurement* content domain

Box 3.5 An example PISA *change and relationships* item classified as *measurement* against TIMSS content domains

INTERNET RELAY CHAT

Mark (from Sydney, Australia) and Hans (from Berlin, Germany) often communicate with each other using "chat" on the Internet. They have to log on to the Internet at the same time to be able to "chat".

To find a suitable time to "chat", Mark looked up a chart of world times and found the following:



Question 1: INTERNET RELAY CHAT

M402Q01 - 0 1 9

At 7:00 PM in Sydney, what time is it in Berlin?

Question 2: INTERNET RELAY CHAT

M402Q02 - 0 1 9

Mark and Hans are not able to chat between 9:00 AM and 4:30 PM their local time, as they have to go to school. Also, from 11:00 PM till 7:00 AM their local time they won't be able to chat because they will be sleeping.

When would be a good time for Mark and Hans to chat? Write the local times in the table.

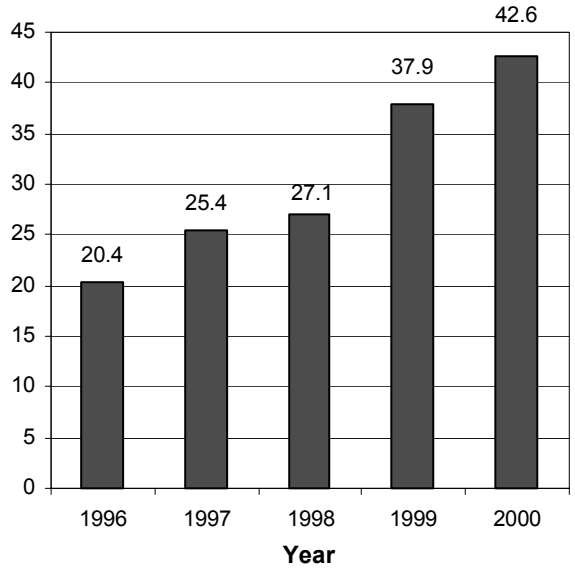
Place	Time
Sydney	
Berlin	

Uncertainty

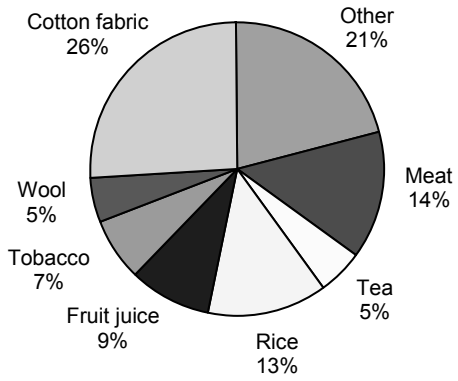
136. This *overarching idea* can be linked to TIMSS' *data* strand. Interestingly, while this strand covers chance and data, TIMSS chooses to label it *data*, and PISA chooses to label it *uncertainty* (chance). It might be less confusing if both surveys just use the label *chance and data* as for traditional curriculum mathematics strand. Of the 20 items classified as *uncertainty* in PISA, 15 are classified as *data*, but five are classified as *number*. These five items involve computing averages and percentages, which could be regarded as *number* or *data* (statistics). Box 3.6 shows a PISA *uncertainty* item classified as *number* (topic *percentages*) against TIMSS content domains.

Box 3.6 An example PISA *uncertainty* item classified as *number* against TIMSS content domains

Total annual exports from Zedland in millions of zeds, 1996-2000



Distribution of exports from Zedland in 2000



EXPORTS

The graphics below show information about exports from Zedland, a country that uses zeds as its currency.

Question 2: EXPORTS

M438Q02

What was the value of fruit juice exported from Zedland in 2000?

137. The last column in Table 3.6 shows the number of PISA items as classified by TIMSS content domains. Notably there are few PISA items in the *algebra* domain, and a relatively large number of

items in the *number* and *data* domains. This is not surprising, given that PISA defines the assessment through an examination of the range of quantitative reasoning used by citizens in everyday life in situations such as "shopping, travelling, cooking, dealing with personal finances, judging political issues, etc." (p.24, OECD, 2003). Such a list of situations/contexts appears to preclude applications of advanced mathematics in very specialised fields, such as the use of transformational geometry in animation, solving differential equations in engineering, forecasting trends and building mathematical models using calculus. While these applications may well be beyond that expected of 15-year-olds and do not appear in either PISA or TIMSS, the underlying mathematical concepts stem from algebra. That is, while ordinary citizens are not required to know a great deal of algebra to become mathematically literate citizens, specialists in mathematics do need to have knowledge in algebra. Algebra is needed for the advance of the modern world, but needed only by a few specialists who can work with technological developments This seems the key distinction between the orientations of the PISA and TIMSS frameworks, where PISA focuses on everyday needs of citizens in terms of using mathematics, while TIMSS focuses on mathematics as a discipline to be used for potential applications in all fields.

138. Finally, with regard to mathematics content classifications, PISA does not provide further breakdown beyond the overarching ideas. This is a mechanism to avoid testing fragments of skills. While this reflects the emphasis on literacy in PISA, for the purposes of this report such a broad classification scheme poses a potential problem. The classification of a PISA item into one traditional mathematics content domain is a matter of judgement, as the framework is not designed with this in mind. In particular it is challenging to map items in the *change and relationships* overarching idea to traditional content domains. Consequently, it is acknowledged that the classification of PISA test items by traditional mathematics content domains could vary considerably.

Number of TIMSS items by PISA overarching ideas

139. A further comparison between PISA and TIMSS frameworks is illustrated by a re-classification of TIMSS items by PISA overarching ideas. Table 3.7 shows the distribution of TIMSS items cross-classified by TIMSS content domains and PISA overarching ideas.

Table 3.7 Tally of TIMSS items classified by PISA overarching ideas and TIMSS content domains

		PISA overarching ideas				
		Quantity	**Space and shape**	**Change and relationships**	**Uncertainty**	**Total**
TIMSS content domains	**Number**	51	3	3		**57**
	Algebra	13	1	33		**47**
	Measurement	15	15	1		**31**
	Geometry		30	1		**31**
	Data	1		10	17	**28**
	Total	**80**	**49**	**48**	**17**	**194**

Table 3.7 shows that PISA overarching idea *Quantity* is closely related to TIMSS *Number* content domain, *Space and Shape* to TIMSS *Geometry and Measurement* content domains, *Change and relationships* to TIMSS *Algebra* content domain, and *Uncertainty* to TIMSS *Data* content domain. In addition, the TIMSS test does not have many items classified as PISA *Uncertainty* overarching idea.

A comparison of PISA's processes dimension with TIMSS' cognitive domains

140. The descriptions for PISA's processes dimension are not dissimilar to those for TIMSS cognitive domains. Both are concerned with cognitive demands (other than mathematics content) in the process of

solving a mathematics problem. Below are some comparisons between the three PISA competency clusters and the four TIMSS cognitive domains.

141. The PISA *reproduction* cluster can be linked to TIMSS cognitive domain *knowing facts and procedures*, and perhaps also covers some of the skills listed in TIMSS *using concepts* domain. The PISA *connections* cluster can be related to TIMSS *using concepts* and *solving routine problems* domains. The PISA *reflection* cluster can be linked to TIMSS *reasoning* domain, where non-routine problems are presented to students.

142. To facilitate comparisons between PISA competency clusters and TIMSS cognitive domains, PISA items were classified according to TIMSS cognitive domains. Table 3.8 shows a tally of the PISA items cross-classified according to PISA competency clusters and TIMSS cognitive domains.

Table 3.8 Tally of PISA items classified by PISA competency clusters and TIMSS cognitive domains

		PISA competency clusters			
		Reproduction	**Connections**	**Reflection**	**Total**
TIMSS cognitive domains	**Knowing facts and procedures**	15	8	0	**23**
	Using concepts	5	8	1	**14**
	Solving routine problems	4	8	3	**15**
	Reasoning	2	16	15	**33**
	Total	**26**	**40**	**19**	**85**

143. As expected, most of the items classified as PISA *reproduction* items are classified as TIMSS *knowing facts and procedures* items. And most of the items classified as PISA *reflection* items are classified as TIMSS *reasoning* items. However, for items classified as PISA *connections* items, there is a spread across the TIMSS classifications, but with more items in the TIMSS *reasoning* domain. Overall, the proportions of PISA items classified by TIMSS cognitive domains are quite different from the proportions of TIMSS items in the cognitive domains (compare with Table 3.2) where the most number of TIMSS items are in the *solving routine problems* domain, and the most number of PISA items are in the *reasoning* domain. It appears that PISA has succeeded in moving a little away from routine problem solving, and moving towards assessing students' ability to solve non-routine problems. However, about one quarter of the items are still in the *knowing facts and procedures* domain, where students are assessed on recall and applications of basic procedures. The inclusion of these lower level items is necessary, otherwise students at lower levels on the mathematical proficiency scale will not find PISA tests accessible. Nevertheless, the differences between PISA and TIMSS in the distributions of items across the cognitive domains reflect the different approaches to the development of the frameworks.

Characteristics of tests and items

144. In this section, item features, such as item format, unit structure, and amount of reading involved, are compared between PISA and TIMSS.

Item Format

145. Both PISA and TIMSS use multiple-choice and constructed-response item formats, as both surveys recognise that a multiple-choice format is not suited for students to demonstrate their abilities to communicate solutions, make interpretations, construct models and perform other more complex tasks. On the other hand, the cost of scoring constructed-response items can become prohibitively expensive in

large-scale assessments, so the objectively scored multiple-choice item format is also used. Table 3.9 shows the proportions of items in multiple-choice or constructed-response format for the two surveys.

Table 3.9 Proportions of items by item format

	PISA	TIMSS
Multiple-choice	33%	66%
Constructed-response	67%	34%

146. Table 3.9 shows that PISA has far more items in constructed-response format than TIMSS. In fact, two thirds of the items in PISA are in constructed-response format, while only one third of the items in TIMSS are in constructed-response format.

147. In general, constructed-response items are more discriminating than multiple-choice items (*e.g.*, Routitsky and Turner, 2003), so one would expect PISA tests to show higher test reliability than TIMSS if the same number of items is administered. However, TIMSS tests have a higher test reliability than PISA tests (see the section *Amount of assessment material* in Chapter 2), as more items are administered in TIMSS, on average, to each student, even though the administration time is shorter. That is, TIMSS items tend to be shorter items, while PISA items require more time to answer. For example, PISA items have more words in the questions, and will require more time to process the information. The following section compares the amount of reading in PISA and TIMSS.

Amount of reading

148. PISA items involve more reading than TIMSS items. To convey real-world problems situations, more words are required to explain about the problem setting, the constraints, and other parameters that will need to be assumed. A sample of items was randomly selected from PISA and TIMSS, and the number of words in the stem of each item is recorded, as shown in Table 3.10.

Table 3.10 Number of words in item stem in randomly selected PISA and TIMSS items

TIMSS cluster M02 Question number	Number of words in item stem	PISA booklet 3 Question number	Number of words in item stem
1	19	1	55
2	25	2	59
3	32	3	33
4	20	4	72
5	23	5	30
6	34	6	130
7	5	7	78
8	9	8	33
9	55	9	78
10	35	10	12
11	18	11	34
12	18	12	50
13	11	13	53
14	6	14	23
15	22	15	101
Mean	**22**	**Mean**	**56**
Standard deviation	13	Standard deviation	32
Standard error	3.4	Standard error	7.9

149. From a random sample of 15 items in TIMSS and 15 items in PISA, the average number of words in a PISA item stem is about twice as many as the average number of words in a TIMSS item stem.

That is, the reading load is considerably heavier in PISA than in TIMSS. In addition, PISA items require more interpretation of the problem statements over and above the straightforward reading of words.

Unit structure

150. Given the amount of texts required in PISA to explain problem settings and contexts, it is often more efficient to ask more than one question within a problem setting, to reduce the amount of reading for each individual task. In cases where more than one question is asked in relation to one stimulus material, the group of questions is referred to as a unit. In PISA, 41% of the items are stand-alone items, while the rest are grouped in units. In TIMSS, 85% of the items are stand-alone items. The effect of having unit structures instead of stand-alone items is that items within a unit are typically more similar to each other, thus violating the local independence assumption of items under item response modelling. That is, items within a unit could potentially collect the same piece of information, rather than independent pieces of information, about a student's proficiency being assessed. This, of course, will be the extreme case. In fact, in PISA, care was taken to ensure that the correctness of a response for an item would not depend on the correctness of the answer on another item. But it is possible that students may have some familiarity (or unfamiliarity) with particular contexts and situations, and the correctness of the responses to items within a unit could be a little more similar than they would have been, had the contexts/situations been completely different. Therefore, instead of having collected, say, 20 independent pieces of information, one had only collected 18 pieces of information, Consequently, the reported test reliability could be a little inflated. The achievement scores, however, are not expected to be biased due to the use of units in the assessment, since duplicate information would still be "correct" information.

Summary

151. This chapter compares PISA and TIMSS frameworks and notes similarities and differences. The extent of the impact of these differences on achievement results is discussed in Chapter 4. The following is a summary of similarities and differences between the PISA and TIMSS frameworks.

Aspects of the Frameworks	Similarities	Differences
Approach	Both PISA and TIMSS framework development involved extensive consultative processes with participating countries and mathematics education experts.	PISA framework is primarily expert driven with endorsements by countries. TIMSS framework is primarily country driven with endorsements by experts.
Organising principles	Both PISA and TIMSS organise the framework with content and cognitive dimensions	
Content organisation		TIMSS adopts traditional mathematics content domains: *Number*, *algebra*, *measurement*, *geometry* and *data*. PISA uses phenomenological approach to categorise problems based on the kinds of applications of mathematics. Four overarching ideas are identified: *quantity, space and shape, change* and *relationships, uncertainty*.
Content balance		TIMSS covers a wider range of curriculum contents than PISA. PISA has few items in *algebra*, *measurement* and *geometry*, but more items in *number* and *data*.
Cognitive dimension	While the labels are different for areas of the cognitive dimension, both PISA and TIMSS describe cognitive processes (or competencies) in terms of progressions from simple to complex tasks, with *knowing facts/reproduction* at the lowest level, and *reasoning/reflection* at the highest level.	
Item format		Two-thirds of the items in TIMSS are of multiple-choice format, while only one-third of the items in PISA are multipl-choice items
Amount of reading		On average, PISA items have around twice as many words as TIMSS items.

CHAPTER 4 - COMPARISON OF PISA AND TIMSS ACHIEVEMENT RESULTS

152. This chapter compares PISA and TIMSS mathematics achievement scores for countries participating in both surveys, and identifies factors that are associated with the observed differences in results.

Comparisons of country mean scores

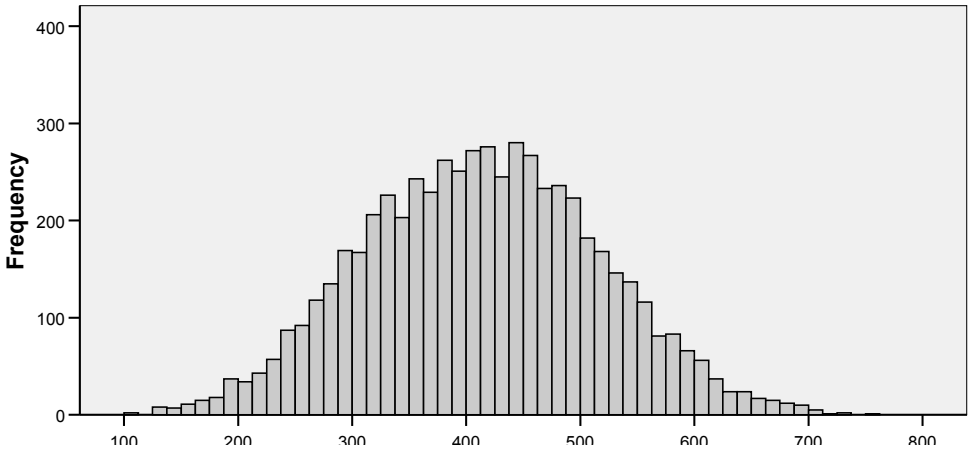
153. In reporting student scores, both PISA and TIMSS transformed students' IRT scores into a metric that had a mean of 500 and a standard deviation of 100 based on a specified reference population. In PISA, for mathematics the reference population was the group of OECD countries in the 2003 survey, while in TIMSS, the reference population was the group of participating countries in 1995 Grade 8 TIMSS mathematics survey (this was to ensure comparability of results to the 1995 data). Consequently, PISA and TIMSS do not have strictly aligned definitions for central location and spread for the distributions of reported scores, given that the reference populations are different (See Box 4.1 for explanations of central location and spread of distributions). That is, a value of 500 in PISA is not directly comparable to a value of 500 in TIMSS, since 500 is the mean for a different set of countries in PISA than in TIMSS. Further, one PISA score point is not the same as one TIMSS score point in representing achievement differences, since the scores have been multiplied by a factor proportional to the standard deviation of the respective scores distribution. Consequently, a direct comparison of the reported scores of countries in TIMSS and PISA is only valid for comparing rankings of countries, but not how far apart the countries are from each other.

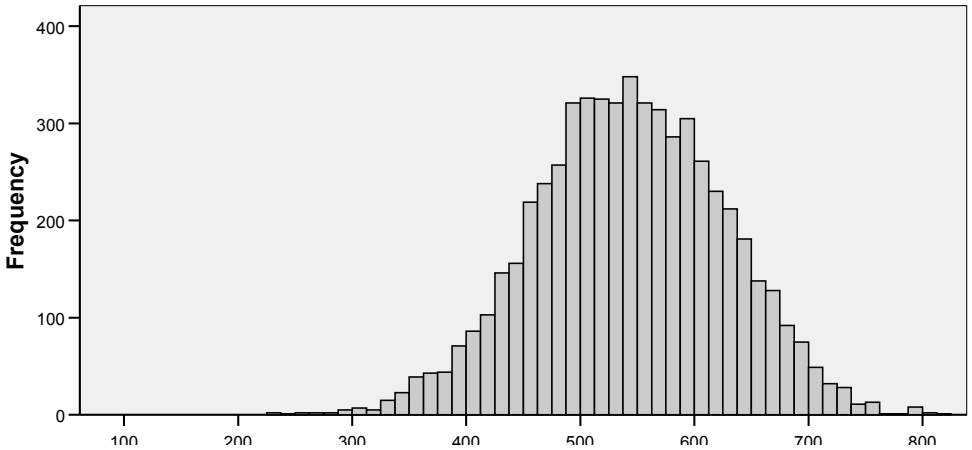
154. The reported PISA and TIMSS country mean scores for the 22 countries/regions that participated in both PISA and TIMSS are shown in Table 4.1 in columns 2 and 5, with associated standard errors for the means in columns 3 and 6. It is important to note that data for England are included in this report for illustration, but that England did not meet the required response rates and therefore did not satisfy the technical requirements to confidently say that results were comparable internationally in both the TIMSS and PISA 2003 surveys. For example, the country average for England was not published in the release of the PISA 2003 initial results (OECD, 2004). All achievement results for England should therefore be interpreted with caution.

Box 4.1 Central Location and Spread of Distributions

The following are two example distributions of mathematics scores for two groups of students. It can be seen that the performance of the first group is lower than that of the second group, in that the distribution of the first group is further to the left of the scale (*i.e.*, lower scores). To describe the general "location" of a distribution, statistics such as mean and median are useful. These statistics are referred to as statistics for central tendency. They provide information about where the "centre" of the distribution is located.

Further, it can be seen that the distribution for the first group is more spread out than that for the second group. That is, the range of scores for the first group is wider. Statistics such as range and variance are useful to describe the spread of a distribution.





155. The countries in Table 4.1 are arranged in decreasing order of PISA country means (column 2). A glance down the column of TIMSS country means (column 5) shows that these are also in approximately decreasing order, with some disordering. For example, while the Flemish Community of Belgium has a

similar mean score to Hong Kong-China in PISA, there is a large difference in TIMSS scores between the two regions. The correlation between PISA and TIMSS country mean scores is 0.84, showing that, in general, there is a reasonable agreement between the results of the two surveys. Note that Indonesia and Tunisia have mean scores much lower than those of the other countries. If these two countries are omitted in the table, the correlation between the remaining 20 countries is 0.66, which still indicates an association between the PISA and TIMSS scores, although the relationship is not extremely strong.

156. To facilitate comparisons between PISA and TIMSS scores, two sets of *standardised* scores were computed. The PISA country mean scores were standardised to have a mean of zero and a standard deviation of one (column 4) for the 22 countries. Similarly, TIMSS country mean scores were standardised to have a mean of zero and a standard deviation of one (column 7) for the same set of countries. That is, a value of 1.01 for Hong Kong-China in PISA indicates that the mean PISA score for Hong Kong-China is 1.01 PISA standard deviations away from the average of the 22 country means. Similarly, a value of 1.66 for Hong Kong-China in TIMSS indicates that the mean TIMSS score for Hong Kong-China is 1.66 TIMSS standard deviations away from the mean of the 22 countries. That is, the TIMSS score for Hong Kong-China is actually relatively higher than the PISA score when measured in terms of the "distances" from the other countries under comparison.

157. These standardised scores are now comparable between PISA and TIMSS. Had countries performed in the same way in PISA and TIMSS, one would expect the standardised PISA and TIMSS scores to be very similar for each country. If a country has very different standardised scores in PISA and in TIMSS, then one might conclude that the country performed differently in PISA and in TIMSS.

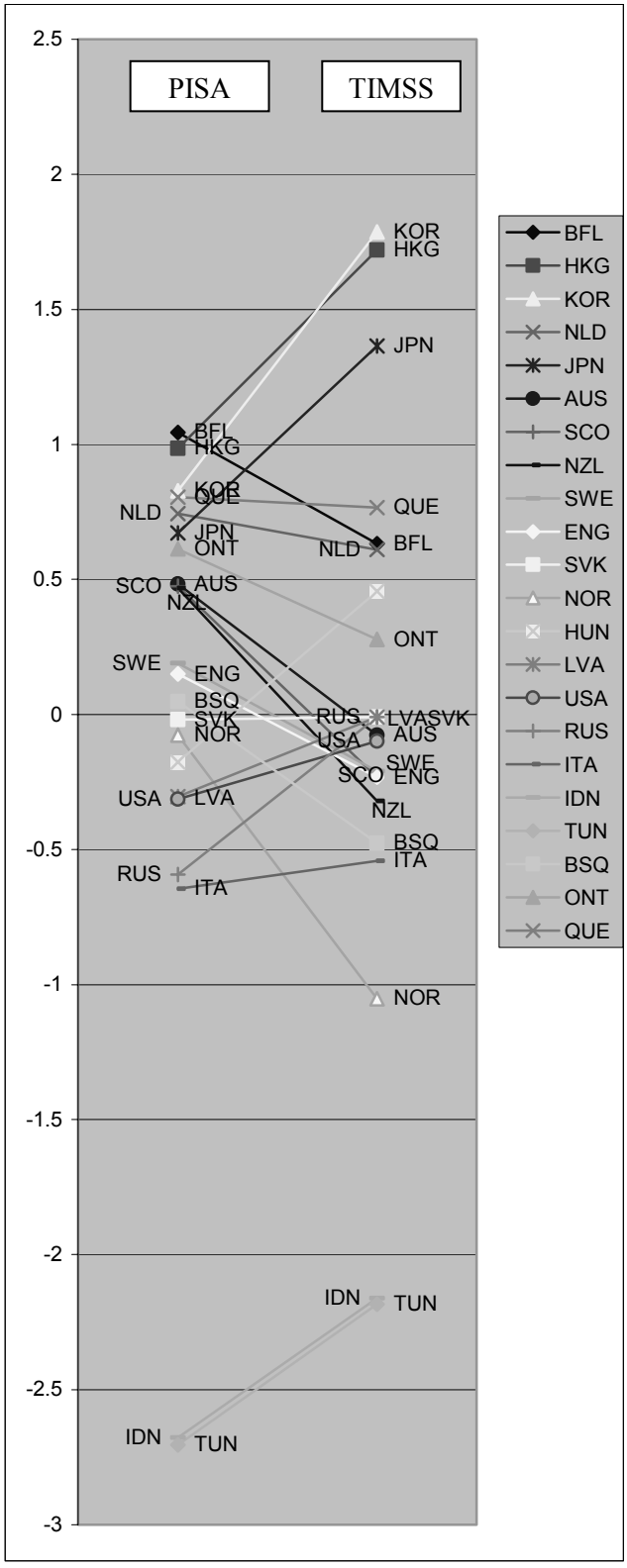
Table 4.1 PISA and TIMSS country mean scores for countries participating in both Surveys in 2003

	PISA			TIMSS		
	Country mean	Standard error	Standardised score	Country mean	Standard error	Standardised score
Belgium (Fl.)	553	(2.1)	1.04	537	(2.8)	0.63
Hong Kong-China	550	(4.5)	0.99	586	(3.3)	1.72
Korea	542	(3.2)	0.83	589	(2.2)	1.79
Quebec, Canada	541	(5.0)	0.81	543	(3.0)	0.77
Netherlands	538	(3.1)	0.74	536	(3.8)	0.61
Japan	534	(4.0)	0.67	570	(2.1)	1.37
Ontario, Canada	531	(3.5)	0.61	521	(3.1)	0.28
Australia	524	(2.1)	0.48	505	(4.6)	-0.08
Scotland	524	(2.3)	0.47	498	(3.7)	-0.23
New Zealand	523	(2.3)	0.47	494	(5.3)	-0.32
Sweden	509	(2.6)	0.19	499	(2.6)	-0.21
England[1]	507	(2.9)	0.15	498	(4.7)	-0.23
Basque country, Spain	502	(2.8)	0.05	487	(2.7)	-0.48
Slovak Republic	498	(3.3)	-0.02	508	(3.3)	-0.01
Norway	495	(2.4)	-0.08	461	(2.5)	-1.05
Hungary	490	(2.8)	-0.18	529	(3.2)	0.46
Latvia	483	(3.7)	-0.30	508	(3.2)	-0.01
United States	483	(2.9)	-0.31	504	(3.3)	-0.10
Russian Federation	468	(4.2)	-0.59	508	(3.7)	-0.01
Italy	466	(3.1)	-0.65	484	(3.2)	-0.54
Indonesia	360	(3.9)	-2.68	411	(4.8)	-2.16
Tunisia	359	(2.5)	-2.70	410	(2.2)	-2.18
Average	499			508		0.00
Standard deviation	51.9			45.1		1.00

1. England did not achieve the required response rate set by PISA and set by TIMSS for ensuring that a representative sample is drawn for the country. The reliability of the mean scores therefore should be treated with some reservation.

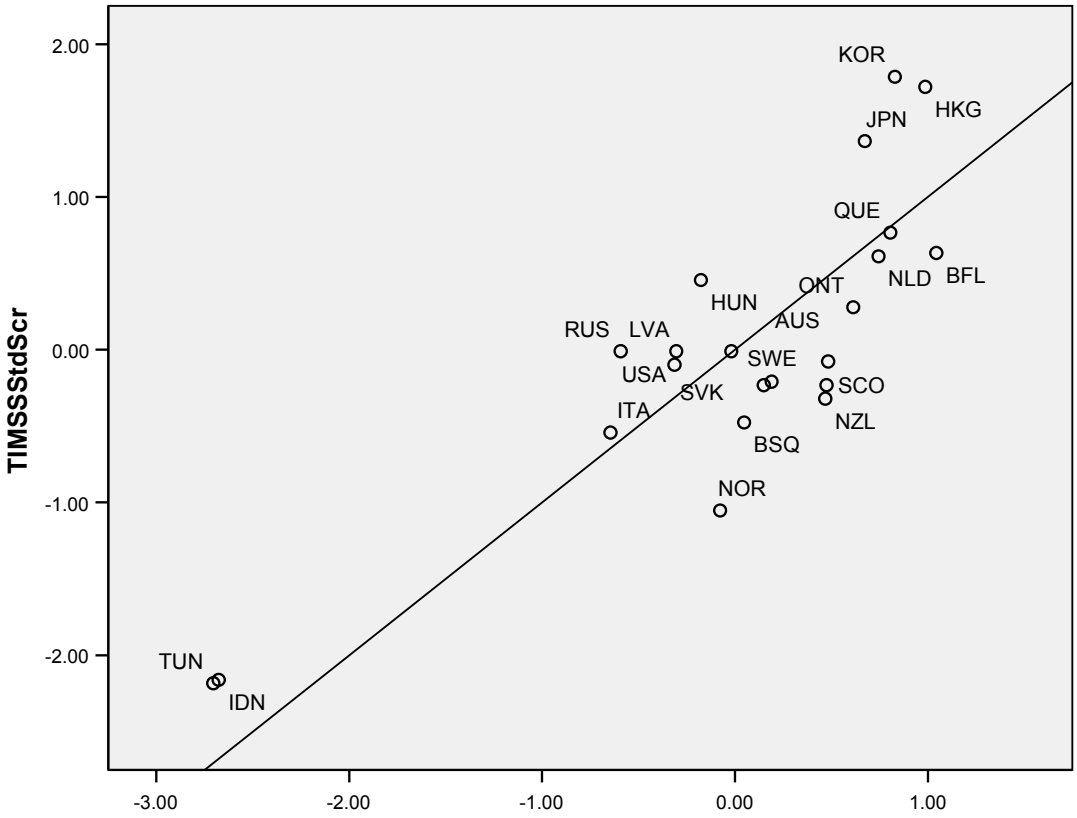
158. The distances between countries in standardised PISA and TIMSS scores for the 22 countries are shown graphically in Figure 4.1

Figure 4.1 Plot of PISA and TIMSS standardised country mean scores



159. A number of observations can be made about Figure 4.1. First, for TIMSS scores, there is a large "gap" between Asian countries and other countries, while, for PISA scores, the "distance" between Asian countries and other countries narrows. This means that there is a greater difference between the group of Asian countries and other countries in performance on the TIMSS test. This difference is not so marked in PISA. Second, there appears to be more separation between countries in achievement levels in PISA, while, in TIMSS, there is a clustering of countries around the overall mean score. Third, the countries that "moved up" from PISA to TIMSS are mostly Asian countries and Eastern European countries, and the countries that moved "down" from PISA to TIMSS tend to be Western countries. As a consequence, most English-speaking countries are ahead of Eastern European countries in PISA, but many of them are behind Eastern Europeans countries in TIMSS. This difference is more clearly shown in Figure 4.2. The countries above the line are those that performed relatively better in TIMSS while those below the line performed relatively better in PISA within the comparison of the 22 countries.

Figure 4.2 Standardised country mean scores: PISA 2003 versus TIMSS 2003



1. England did not meet participation requirements either in TIMSS 2003 or PISA 2003. Mean scores therefore are not comparable to those of other participating countries.

160. Interestingly, the same pattern of differences was observed between PISA 2000 and TIMSS 1999 results (Wu, 2005), where countries with higher standardised PISA mean scores were Australia, Canada, Finland, New Zealand, the United Kingdom and the United States, and countries with higher standardised TIMSS scores were the Czech Republic, Hungary, Italy, Japan, Korea, Hong Kong-China, Latvia and the Russian Federation. The consistent findings for two PISA and TIMSS cycles indicate that the observed pattern is unlikely to be due to chance. There is likely to be systematic differences between the PISA and TIMSS tests in relation to systematic differences in education systems in countries. The following sections present an examination of possible factors to explain the observed performance differences.

Explaining the Differences between PISA and TIMSS Results

161. To identify factors that may explain the observed differences between PISA and TIMSS country rankings as shown in Figure 4.2, test characteristics of PISA and TIMSS are examined, with a focus on those characteristics where PISA and TIMSS differ. These include age/grade sampling method, mathematics content differences and reading demand of test items (see Chapters 2 and 3).

Years of Schooling and Age at time of testing

162. In Chapter 2, age and grade differences between PISA and TIMSS were discussed. TIMSS population definition controls for the number of years of schooling, but the age of students varies more across countries than in PISA. In PISA, the age of sampled students is controlled, but the number of years of schooling varies across countries. Could these explain, at least in part, the observed differences in results between the two surveys? To answer this question, we first examine the inter-relationship between the two variables: *age at time of TIMSS testing*, and *years of schooling at time of PISA testing*. We then relate these variables to achievement results.

163. For each country, the average age at time of TIMSS 2003 testing is given in the TIMSS mathematics report (Exhibit 2, IEA, 2003). Table 4.2 (column 2) shows the average age for Korea and Norway, as an example.

Table 4.2 Relationship between age at time of testing in TIMSS and years of schooling at time of testing in PISA

	Average age at time of TIMSS test (at 8 years of schooling)	Number of years of schooling at time of PISA test (at 15.7 years old)
Korea	14.6	around 9
Norway	13.8	around 10

164. Korea's sample in TIMSS has an average age of 14.6. That is, students with 8 years of schooling in Korea are around 14.6 years of age. Therefore, when students in Korea reach 15.7 years old (the average of the PISA sample), one would expect the students to have 9 years of schooling. Similarly, for Norway, students are 13.8 years of age when they are in Grade 8. One would then expect 15.7 year-olds (the PISA sample) to be in Grade 10. More generally, the number of years of schooling at time of PISA testing can be approximately given by

number of years of schooling at time of PISA testing = (15.7 – *age at time of TIMSS testing*) + 8,

as (15.7 – *age at time of TIMSS testing*) gives the additional number of years of schooling between TIMSS and PISA population definitions, to the 8 years of schooling controlled for in the TIMSS samples. So, in fact, the two variables, *age at time of TIMSS testing*, and *years of schooling at time of PISA testing* are essentially the same variable, as one is a linear transformation of the other:

number of years of schooling at time of PISA testing = 23.7 – *age at time of TIMSS testing* *(1)*

165. To verify this relationship between *age at time of TIMSS testing*, and *years of schooling at time of PISA testing*, an attempt was made to compute the average years of schooling at the time of testing in PISA

using information from the PISA survey. An approximate estimate was made based on the following three pieces of information:

6. The grade variable from the PISA student questionnaire. This variable is meant to provide the number of years of schooling. However, it turned out that this variable alone was too "coarse" for the purpose of estimating years of schooling to the accuracy of fractions of a year.

7. The start of the academic year in each country.

8. The actual testing date of PISA in each country.

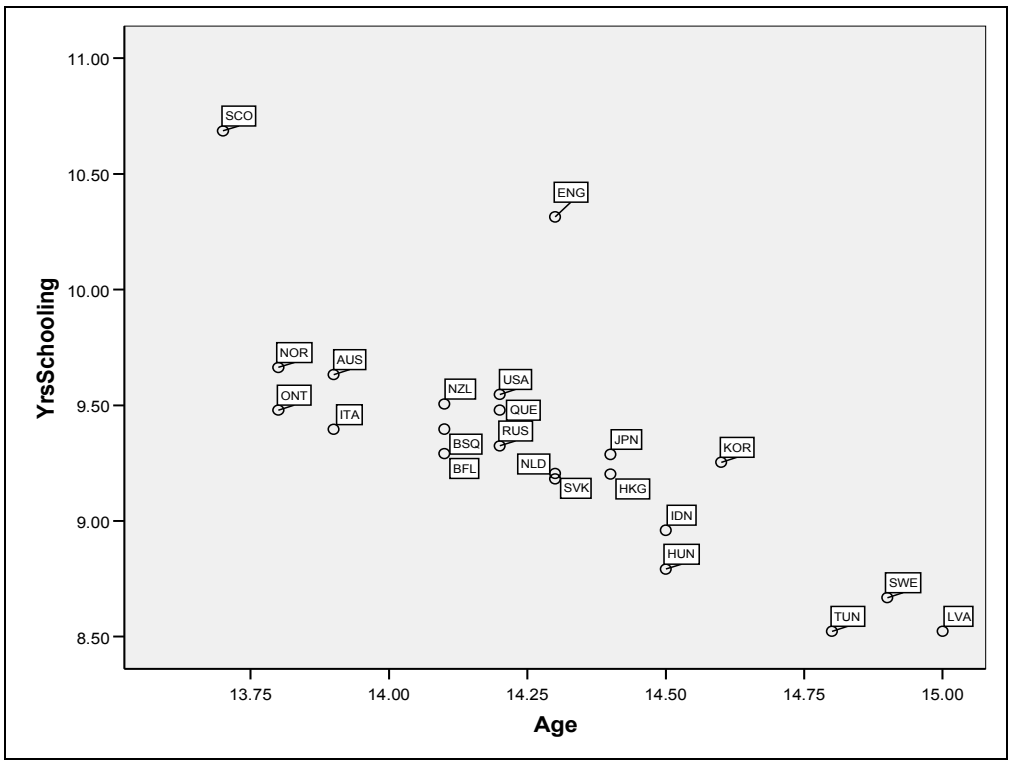
166. Combining (2) and (3), one is able to estimate fractions of a year of schooling of the Grade the students were placed at the time of PISA testing. It should be noted that there is not a great deal of confidence that the grade variable (point 1, above) actually captures the number of years of schooling that is comparable across countries. The estimated number of years of schooling for each country, derived as described, is shown in Table 4.3, as well as the age at the time of testing for TIMSS, for comparison. The entries are arranged in increasing order of the number of years of schooling at the time of testing of PISA, as estimated from PISA data.

Table 4.3 Number of years of schooling at the time of PISA testing (estimated from PISA data) versus Average age at time of TIMSS testing

	Number of years of schooling at the time of PISA testing (estimated from PISA data)	Average age at time of TIMSS testing
Tunisia	8.52	14.8
Latvia	8.52	15
Sweden	8.67	14.9
Hungary	8.79	14.5
Indonesia	8.96	14.5
Netherlands	9.18	14.3
Hong Kong-China	9.20	14.4
Slovak Republic	9.21	14.3
Korea	9.25	14.6
Japan	9.29	14.4
Belgium (Fl.)	9.29	14.1
Russian Federation	9.32	14.2
Italy	9.40	13.9
Spain-Basque	9.40	14.1
Canada-Ontario	9.48	13.8
Canada-Quebec	9.48	14.2
New Zealand	9.51	14.1
United States	9.55	14.2
Australia	9.63	13.9
Norway	9.66	13.8
England	10.31	14.3
Scotland	10.69	13.7

167. Table 4.3 shows that, by and large, PISA students in the Western countries tend to have had a higher number of years of schooling than the PISA students in Asian and Eastern European countries. At the same time, there appears to be a negative relationship between the number of years of schooling at the time of testing of PISA and the age at the time of testing of TIMSS, as one would expect from equation *(1)*. This relationship can be seen more readily in a scatter plot of the two variables, as shown in Figure 4.3.

Figure 4.3 PISA years of schooling versus TIMSS age of testing



168. The correlation between PISA years of schooling (estimated from PISA data) and TIMSS age of testing is -0.77 ($R^2 = 0.59$), showing a strong relationship between these two variables. This is particularly striking given that the variable, number of years of schooling in PISA, was an approximation constructed for this report. Consequently, the age at time of TIMSS testing could be regarded as a proxy variable for years of schooling in PISA, given the theoretical relationship between these two variables as shown in Equation (1), as well as the empirical validation of the relationship as shown in Figure 4.3. England and Scotland appear to be outliers in Figure 4.3. If England and Scotland are removed from Figure 4.3, the correlation between the two variables is -0.9. This strong relationship between the two variables in Figure 4.3 suggests that Grade-based samples in TIMSS have been well controlled for the number of years of schooling.

Box 4.2 The interpretations of *R* and R^2 in regression models

Consider a regression analysis where a dependent variable Y is to be explained, or predicted, by an independent variable, X. The regression equation can be written as

$$Y = a + bX$$

where a is called a regression constant and b is a regression coefficient. a and b are to be estimated in the regression analysis from the (X and Y) data pairs. Typically, regression analysis also reports *R* and R^2. In the case where there is only one explanatory variable (X in the above example), *R* is the correlation coefficient between X and Y. The correlation coefficient, *R*, is a measure of association between two variables, ranging between -1 and 1. If a plot of (X and Y) pairs fall exactly on a straight line with a positive gradient, then the correlation coefficient will be 1. If the line has a negative gradient, then the correlation coefficient will be -1. If there is no association between X and Y, then the correlation coefficient will be zero. The square of the correlation coefficient (R^2) is called the coefficient of determination. R^2 can be shown to be a measure of the proportion of sample variation in the dependent variable Y which is explained by the values of the independent variable X:

$$R^2 = \frac{\text{explained variation in } Y}{\text{total variation in } Y}$$

When there are more than one explanatory (or independent) variable, as shown below:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \cdots$$

R^2 shows the proportion of variation in Y explained by the combined set of the independent variables. In this case, *R* is called the coefficient of multiple correlation, and R^2 is called the coefficient of multiple determination.

Impact of Years of Schooling on Student Performance in Mathematics

169. Table 4.4 shows the list of countries arranged in order of how much better the countries performed in TIMSS than in PISA. In this table, the metric for expressing score differences is in "TIMSS score" unit, and not in standardised mean scores as computed in Table 4.1. That is, PISA scores have been converted to have the same mean and standard deviation of the TIMSS scores for the 22 countries, and then the difference between TIMSS and PISA scores is computed. In this way, the magnitude of the differences can be more easily interpreted than standardised scores, since one can discuss the average gain in TIMSS score unit, with one additional year of schooling, for example. The countries at the top of Table 4.4 performed better in TIMSS than in PISA; the countries at the bottom of the table performed better in PISA. In addition, for each country, the average age of students in the TIMSS[12] assessment is also shown.

12. The average age at time of testing in TIMSS is taken from the TIMSS 2003 International Report (IEA, 2003).

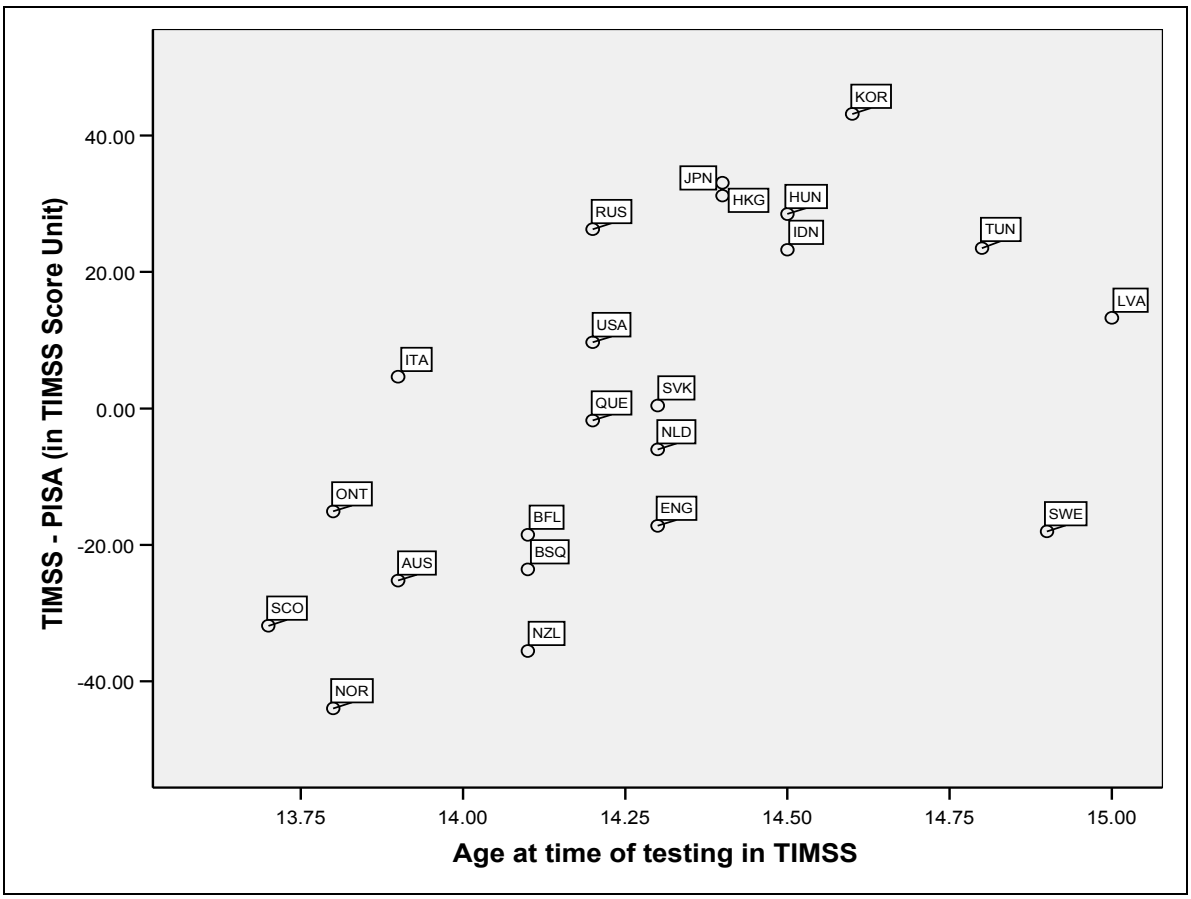
Table 4.4 Comparative performance in PISA and TIMSS, versus age of testing in TIMSS

		Difference in country mean scores (TIMSS - PISA), in "TIMSS score" unit	Average age at time of testing in TIMSS	Number of years of schooling at time of PISA testing (estimated from PISA data)
Better in TIMSS	Korea	**43.16**	14.6	9.25
	Hong Kong-China	**33.08**	14.4	9.20
	Japan	**31.19**	14.4	9.29
	Hungary	**28.50**	14.5	8.79
	Russian Federation	**26.26**	14.2	9.32
	Tunisia	**23.48**	14.8	8.52
	Indonesia	**23.25**	14.5	8.96
	Latvia	**13.26**	15.0	8.52
	United States	**9.69**	14.2	9.55
	Italy	**4.64**	13.9	9.40
	Slovak Republic	**0.40**	14.3	9.21
Better in PISA	Quebec, Canada	**-1.77**	14.2	9.48
	Netherlands	**-6.01**	14.3	9.18
	Ontario, Canada	**-15.09**	13.8	9.48
	England[1]	**-17.21**	14.3	10.31
	Sweden	**-18.03**	14.9	8.67
	Belgium (Fl.)	**-18.53**	14.1	9.29
	Basque country, Spain	**-23.59**	14.1	9.40
	Australia	**-25.24**	13.9	9.63
	Scotland	**-31.86**	13.7	10.69
	New Zealand	**-35.57**	14.1	9.51
	Norway	**-43.99**	13.8	9.66

1. PISA results for England are not comparable and therefore this score point difference should be interpreted with caution.

170. Table 4.4 shows that the Asian and Eastern European countries tend to perform better in TIMSS than in PISA. But it also appears that the Asian and Eastern European countries have a slightly older cohort in the TIMSS assessment. The relationship between *differential performance in TIMSS and PISA,* and *age at time of testing in TIMSS* can be better seen from a scatter plot of the two variables, as shown in Figure 4.4, where the vertical scale shows the difference between TIMSS and PISA score (Column 3 of Table 4.4).

Figure 4.4 Relationship between performance in TIMSS and PISA and age at time of testing in TIMSS



171. Apart from Sweden, Figure 4.4 shows a positive relationship between (TIMSS – PISA) and the age at time of testing in TIMSS. The correlation between the two variables is 0.58 (R^2=0.34). This is significantly different from zero with p=0.004. If Sweden is removed from the data set, the correlation between the two variables is 0.713 (R^2=0.51), significantly different from zero with p=0.0003. That is, countries with an older cohort at the time of testing in TIMSS tend to perform better in TIMSS than in PISA. In contrast, Norway and Scotland performed a great deal better in PISA than in TIMSS (see Figure 4.1), and it appears that these two countries/regions have the youngest cohorts (13.7 and 13.8 years old respectively) among the 22 TIMSS countries.

172. Of all 50 participating countries in TIMSS 2003 Grade 8 cohort[13], Sweden is the only Western country with an older cohort (14.9 years old), while all other Western countries have an average age less than 14.3 years. The case for Sweden is discussed further in this chapter when both TIMSS and PISA advantage indices are presented and an explanation is provided for why Sweden appears as an outlier in Figure 4.4.

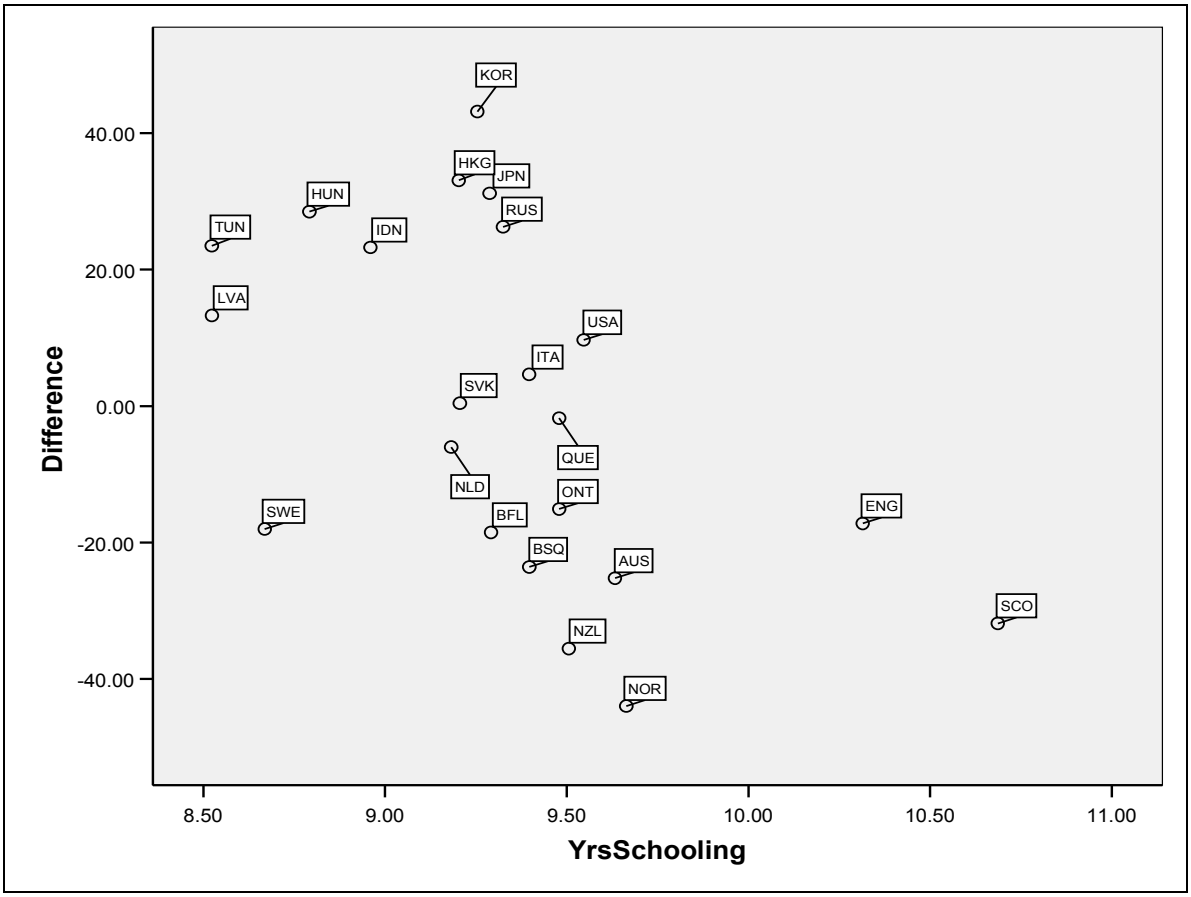
173. Overall, there is a relationship between the differential performance in TIMSS and PISA, and the age of the cohort in the TIMSS sample. It should be remembered, however, that the age of the cohort in the

13 There 46 countries and four benchmarking participants in the TIMSS 2003 Grade 8 cohort.

TIMSS sample is also a proxy for the number of years of schooling at time of PISA testing, as discussed in the previous sections of this chapter.

174. A plot of years of schooling in PISA (as estimated from the PISA data) and differential performance between TIMSS and PISA is shown in Figure 4.5.

Figure 4.5 Relationship between years of schooling in PISA and differential performance between TIMSS and PISA



175. Figure 4.5 shows a negative relationship between the number of years of schooling in PISA and the differential performance between TIMSS and PISA, where the correlation between the two variables is -0.52 (R^2=0.27). This relationship is not as strong as that between the age at the time of testing in TIMSS and the differential performance between TIMSS and PISA, possibly due to the fact that the estimated number of years of schooling in PISA had to be constructed indirectly from a number of data sources. Nevertheless, Figure 4.5 still shows an association between the two variables.

176. The relationships shown in Figures 4.3, 4.4 and 4.5 suggest that the difference between performance in TIMSS and PISA could be related to the age of testing in TIMSS or the number of years of schooling in PISA. For future cycles of testing, it will be helpful to capture more reliable information on the number of years of schooling at the time of testing, to help to understand and interpret the achievement results.

The impact of differences in content balance between PISA and TIMSS

177. From a number of studies (*e.g.*, Routitsky and Zammit, 2002; Zabulionis, 2001) carried out on the comparisons of mathematics education in the international context, it is not surprising to hypothesise that differences in content balance in PISA and TIMSS may lead to the observed differences in country performances as shown in Figures 4.1 and 4.2. In particular, there are different mathematics traditions between countries. The Asian and Eastern European countries stress formal mathematics, while the Western countries place an emphasis on problem solving and application skills (*e.g.*, Leung, Graf & Lopez-Real, 2006). Since the TIMSS assessment is more curricular focused and the PISA assessment is more problem oriented, this could explain the observed differences in the results of the two assessments.

178. In making comparisons between the content of the PISA and TIMSS assessments, PISA items were classified according to the TIMSS content domain classifications (see Table 3.6). The decision to re-classify PISA items according to TIMSS classifications rather than the other way round is because national curricula are mostly structured by traditional mathematics content areas. It will be easier to compare proportions of items in different traditional mathematics content areas in PISA and TIMSS to those in the national curricula.

179. A comparison of the proportions of PISA and TIMSS items by TIMSS content domain classifications is given in Table 4.5.

Table 4.5 Number and proportions of items in PISA and TIMSS by content domains

	PISA		TIMSS		Differences in percentages between PISA and TIMSS
Number	32	38%	57	30%	8%
Algebra	7	8%	47	24%	-16%
Measurement	8	9%	31	16%	-7%
Geometry	12	14%	31	16%	-2%
Data	26	31%	28	14%	17%
Total	**85**	**100%**	**194**	**100%**	

180. It can be seen from Table 4.5 that, in PISA, there are more items in the content domains of *number* and *data*, while, in TIMSS, the proportions of items in each content domain are more evenly spread. There are fewer *algebra* and *measurement* items in PISA when compared to item proportions in TIMSS. Given the "literacy" orientation of the PISA assessment, the distribution of PISA items across the content domains is not surprising. If one surveys the mathematics that most people have to deal with in everyday life, one does not come across solving equations very often. Instead, in everyday life, one needs to interpret information in tables or graphs, or calculate prices/discounts, so that there is a predominance of *number* and *data* applications that one has to deal with. This is not to say that *algebra* is not an important part of mathematics. Algebra is very important for many applications in the technological world. But, by and large, only a small proportion of people specialise in these applications.

Achievement by content domains

181. It will be of interest to examine whether there are differences across countries in achievement scores by content domains. TIMSS 2003 International Mathematics Report published averaged scaled scores by mathematics content areas by country (see IEA, 2003, Exhibit 3.1). The following is an extract from the TIMSS report for the 22 countries that participated in both PISA and TIMSS.

Table 4.6 TIMSS achievement scores by content areas

	Number	Algebra	Measurement	Geometry	Data
Australia	498	499	511	491	531
Belgium (Fl.)	539	523	535	527	546
England[1]	485	492	505	492	535
Canada-Ontario	516	515	520	513	538
Canada-Quebec	546	529	541	542	544
Hong Kong-China	586	580	584	588	566
Hungary	529	534	525	515	526
Indonesia	421	418	394	413	418
Italy	480	477	500	469	490
Japan	557	568	559	587	573
Korea	586	597	577	598	569
Latvia	507	508	500	515	506
Netherlands	539	514	549	513	560
New Zealand	481	490	500	488	526
Norway	456	428	481	461	498
Russian Federation	505	516	507	515	484
Scotland	484	488	508	491	531
Slovak Republic	514	505	508	501	495
Spain-Basque	490	490	488	456	499
Sweden	496	480	512	467	539
Tunisia	419	405	407	427	387
United States	508	510	495	472	527

1. England did not meet participation requirements in TIMSS 2003. Mean scores are therefore not comparable to other countries.

182. Table 4.6 shows that the patterns of performance across mathematics content areas are quite different between countries. For example, in Australia, the average score for *data* is 32 points higher than for *algebra*[14]. In contrast, in the Russian Federation, it is just the opposite, where the score for *algebra* is 32 points higher than the score for *data*. Such differential patterns of content area achievement scores across countries will lead to different aggregate scores if proportions of items from different content areas change. As described in the previous section, PISA and TIMSS have quite different proportions of items from each content area, as shown in Table 4.5.

183. To assess the impact on mathematics achievement score when proportions of items from different content areas change, indices of "PISA advantage" and "TIMSS advantage" were constructed. The methodology is presented below and is based only on TIMSS achievement scores and not PISA achievement scores. Also, the mapping of PISA items to TIMSS content domains that was presented in Chapter 3 is used and the reader is reminded that no definitive categorisation of PISA items into TIMSS content domains is possible. With these caveats in mind the methodology, using Australia as an example, is as follows:

184. The TIMSS average scores by mathematics content areas for Australia are given below:

14 The standard errors of these estimates are around 2 to 5, so a mean difference of 32 is certainly significant.

	TIMSS mean score by content area					Mean of the five content areas[15]
	Number	Algebra	Measurement	Geometry	Data	
Australia	498	499	511	491	531	506

185. The final column shows the mean of the five content area scores (506). The difference between each content area score and the mean of the five content areas is then calculated, as shown below:

	Deviation of content area score from the mean of the five content areas (506)				
	Number	Algebra	Measurement	Geometry	Data
Australia	-8	-7	5	-15	25

186. The deviation from each content area is then weighted by the proportion of items in each assessment. For example, for PISA, the proportions of items are as follows:

PISA item distribution by content areas				
Number	Algebra	Measurement	Geometry	Data
38%	8%	9%	14%	31%

187. The PISA advantage index is computed as the weighted sum of the deviations of content area scores from the mean of the five content areas, weighted by item proportions in PISA. For example, the PISA advantage index for Australia is:

$$(-8) \times 0.38 + (-7) \times 0.08 + 5 \times 0.09 + (-15) \times 0.14 + 25 \times 0.31 = 2.42$$

188. The same method is applied to calculate the TIMSS advantage index, using the proportion of TIMSS items listed below:

TIMSS item distribution by content areas				
Number	Algebra	Measurement	Geometry	Data
30%	24%	16%	16%	14%

189. The TIMSS advantage index for Australia is:

$$(-8) \times 0.30 + (-7) \times 0.24 + 5 \times 0.16 + (-15) \times 0.16 + 25 \times 0.14 = -2.18$$

190. An overall Content Advantage Index is computed as the difference between PISA advantage index and TIMSS advantage index. So, for Australia, the Content Advantage Index is given by:

$$2.42 - (-2.18) = 4.6$$

15 Note that this is the mean of the scores for the five content areas. This mean will be close to the average scaled score for overall mathematics achievement, but not necessarily exactly the same.

191. The Content Advantage Index shows how much more advantage a country has in PISA as compared to in TIMSS. Thus, from the above computations, the PISA assessment gives Australia more "advantage" (4.6 units), than the TIMSS assessment in that Australia is likely to score higher in PISA than in TIMSS, given the composition of items from different content areas in PISA and in TIMSS.

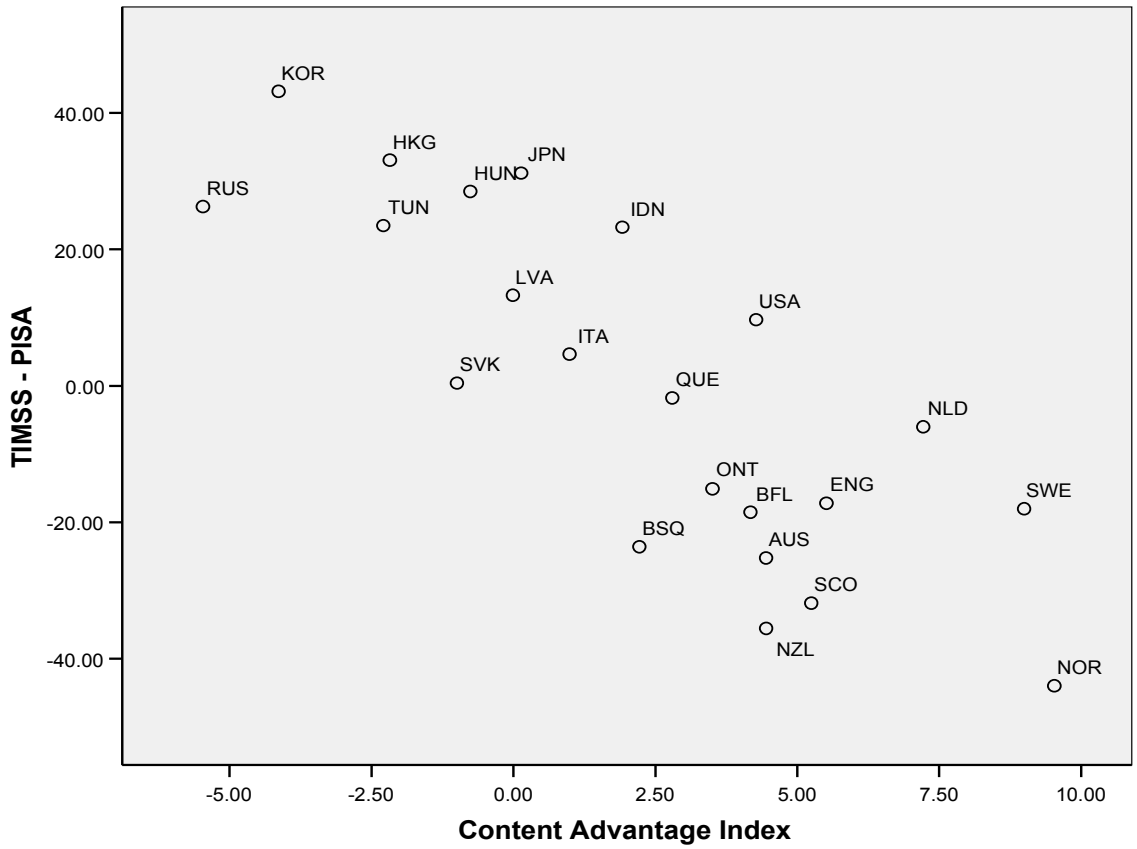
192. The PISA and TIMSS advantage indices, and the Content Advantage Indices were computed for all 22 countries, and these are shown in Table 4.7, together with the difference in TIMSS and PISA achievement mean scores (in TIMSS score unit), arranged in order of the difference in performance in TIMSS and PISA.

Table 4.7 Indices of PISA and TIMSS advantage by country

		Difference in country mean scores (TIMSS - PISA), in "TIMSS score" unit	PISA advantage index	TIMSS advantage index	Content advantage index (PISA adv – TIMSS adv)
Better in TIMSS	Korea	43.16	-2.85	1.29	-4.14
	Hong Kong	33.08	-1.32	0.86	-2.18
	Japan	31.19	-1.58	-1.71	0.14
	Hungary	28.50	0.34	1.10	-0.76
	Russian Federation	26.26	-4.32	1.15	-5.47
	Tunisia	23.48	-0.94	1.35	-2.29
	Indonesia	23.25	3.36	1.45	1.92
	Latvia	13.26	0.05	0.06	-0.01
	United States	9.69	5.27	1.00	4.27
	Italy	4.64	-0.06	-1.05	0.99
	Slovak Republic	0.40	0.45	1.44	-0.99
Better in PISA	Canada-Quebec	-1.77	2.55	-0.25	2.80
	Netherlands	-6.01	5.64	-1.58	7.22
	Canada-Ontario	-15.09	2.20	-1.31	3.51
	England	-17.21	1.94	-3.57	5.51
	Sweden	-18.03	6.45	-2.55	8.99
	Belgium (Fl.)	-18.53	3.75	-0.42	4.18
	Spain-Basque	-23.59	3.16	0.95	2.22
	Australia	-25.24	2.41	-2.04	4.45
	Scotland	-31.86	1.55	-3.69	5.25
	New Zealand	-35.57	1.28	-3.17	4.45
	Norway	-43.99	4.80	-4.73	9.53

193. A simple scatter plot between (TIMSS-PISA) scores (column 3) and the Content Advantage Index (column 6) is given in Figure 4.6.

Figure 4.6 Relationship between difference in TIMSS and PISA scores and the Content Advantage Index



194. Figure 4.6 shows a clear relationship between the difference in TIMSS and PISA mean scores and the content advantage index. When the relative content advantage in PISA is lower than in TIMSS, (TIMSS score – PISA score) tends to be positive, and when the relative content advantage is higher in PISA than in TIMSS, (TIMSS score – PISA score) tends to be negative. The correlation between the two variables in Figure 4.6 is -0.81 (R^2=0.66) with $p<0.001$. Note that, among the 22 countries, Sweden has the second highest content advantage index in favour of PISA. This offsets the fact that students in Sweden have fewer years of schooling at the time of PISA testing (and therefore a higher age cohort in TIMSS).

Predicting PISA Mathematics Country Mean Scores

195. In the previous sections, two factors have been separately identified to have an association with the observed differences between countries' TIMSS and PISA mean scores: *Content Balance* and *Age at time of TIMSS testing (or Years of schooling at time of PISA testing)*. In this section, we explore the combined impact of these two factors on the differences between PISA and TIMSS mean scores. To do this, multiple regressions were carried out. To formulate the regression models, first consider a theoretical relationship:

PISA = TIMSS + (PISA – TIMSS)

where *PISA* denotes a country's PISA mathematics mean score, and *TIMSS* denotes that country's TIMSS mathematics mean score. Consequently, the regression models can be formulated with PISA mathematics country mean score as the dependent variable to be predicted, and TIMSS mathematics country mean score

as the first predictor (independent variable), plus other predictor variables that have a relationship with (PISA – TIMSS) country mean scores.

196. Table 4.8 shows a summary of the multiple regression models used. The simplest model (model 1) is to predict PISA mathematics country mean scores using only TIMSS mathematics country mean scores as a predictor. As discussed at the beginning of this Chapter, the correlation between PISA and TIMSS country mean scores is 0.84. The percentage of variance of PISA mathematics country mean scores explained by the regression model is 71%. That is, around 30% of the variance of PISA country mean scores cannot be explained by their TIMSS country mean scores. Under this regression model, we can compute the predicted PISA mean score for each country and compare it with the observed score as reported in the PISA international report (OECD, 2004). With TIMSS mathematics score as the only predictor, two out of the 22 countries have a predicted PISA mathematics score within the confidence interval of the reported PISA score.

197. For regression model 2, two more variables are added as predictors: *TIMSS Age* and *Content advantage index*. Note that *TIMSS Age* is used as a proxy for *Years of schooling in PISA*. The percentage of variance of PISA mathematics scores explained is 93%, showing a large improvement from regression model 1. Under regression model 2, nine out of the 22 countries have a predicted PISA mathematics score within the confidence interval of the reported PISA score.

198. As there is more reading demand in PISA mathematics items, it is worth exploring whether PISA reading country mean score is a useful predictor for PISA mathematics scores. Regression model 3 explores the relationships between PISA mathematics country mean scores and PISA reading country mean scores. Quite surprisingly, there is a very high correlation (R=0.95) between PISA mathematics country mean scores and PISA reading country mean scores. Under this regression model, five out of 22 countries have a predicted PISA mathematics score within the confidence interval of the reported PISA score.

199. Regression model 4 uses all four predictors: *TIMSS Mathematics country mean score, TIMSS Age, Content advantage index, and PISA Reading score*. Under this model, 97% of the variance of PISA mathematics scores can be explained. Further, 11 out of the 22 countries have a predicted PISA mathematics score within the confidence interval of the reported PISA score. Table 4.9 shows the predicted PISA mathematics scores as compared to the reported scores. It can be seen that, for most countries, the difference between the predicted score and the reported score is less than 10.

200. The four regression models shown in Table 4.8 are not the only models that can be fitted. PISA science country mean scores are also highly correlated with PISA mathematics country mean scores (R=0.97). TIMSS science scores also have a moderately high correlation with PISA mathematics scores (R=0.89). However, the purpose of this section is not so much as to predict PISA mathematics scores per se. The purpose is to illustrate how PISA and TIMSS mathematics scores and other factors are inter-related.

201. It should be noted that different regression models will show different sets of countries as having the best predicted scores. Consequently, Table 4.9 should not be used to make judgements about specific "outlier" countries. If different predictors are used, different countries will be outliers.

Table 4.8 Regression models for predicting PISA mathematics country mean scores

Regression model	To Predict (Dependent variable)	Predictor(s) (Independent variables)	Percentage of variance explained (R^2)	Correlation (R)
1	PISA mathematics	TIMSS Mathematics	71%	0.84
2	PISA mathematics	TIMSS Mathematics TIMSS Age[16] Content advantage index	93%	0.97
3	PISA mathematics	PISA Reading	91%	0.95
4	PISA mathematics	TIMSS Mathematics TIMSS Age Content advantage index PISA Reading	97%	0.99

Table 4.9 Comparisons between Reported and Predicted Country Mean Scores

Country	Reported mean score	Predicted mean score from regression model 4	Difference between reported and predicted scores	Predicted score is within the confidence interval of the reported score
Australia	524	526	-2	✓
Belgium Flemish	553	544	9	
Canada-Ontario,	531	539	-8	
Canada-Quebec	541	541	0	✓
England	507	506	1	✓
Hong Kong - China	550	546	4	✓
Hungary	490	497	-7	
Indonesia	360	371	-11	
Italy	466	479	-13	
Japan	534	533	1	✓
Korea	542	557	-15	
Latvia	483	483	0	✓
New Zealand	523	515	8	
Norway	495	494	1	✓
Russian Federation	468	458	10	
Scotland	524	521	3	✓
Slovak Republic	498	479	19	
Spain-Basque country	502	492	10	
Sweden	509	506	3	✓
The Netherlands	538	534	4	✓
Tunisia	359	355	4	✓
United States	483	502	-19	

16 The variable Age at time of TIMSS test is used as a proxy for Years of schooling at time of PISA test.

Implications of differential performance of countries in content domains

202. The findings from previous sections show that differential performance of countries in content domains can explain, to a large extent, different country rankings in PISA and TIMSS. An implication of this finding is that the reported combined mathematics score must be interpreted in relation to the composition of the test in terms of content balance. For example, if a test consists only of TIMSS Data items, then Sweden ranks 7th out of the 22 countries considered above. If a test consists only of TIMSS Algebra items, then Sweden ranks 18th out of the 22 countries. In contrast, out of 22 countries, Hungary ranks 13th in TIMSS Data content domain, but 4th in TIMSS Algebra content domain. Consequently, tests consisting of different balance of content domains will likely to produce different rankings of countries. Any statement about how a country performed in "mathematics" must be carefully interpreted.

203. One may argue that the performance of countries at the level of content domains may be more informative. As PISA and TIMSS have different content classifications, it will be difficult to cross-check results at the content domain level. Of the five TIMSS content domains and four PISA overarching ideas, the best match is perhaps between TIMSS Data domain and PISA Uncertainty overarching idea (see Chapter 3 for comparisons of the PISA and TIMSS frameworks). The correlation between country mean scores in TIMSS Data content domain and PISA Uncertainty overarching idea is 0.93 for the 22 countries. This is much higher than the correlation between the combined mathematics scores (0.84). This suggests that, if the contents of the tests are aligned, the results will be more similar. For other TIMSS content domains and PISA overarching ideas, it is difficult to form a one-to-one match between TIMSS and PISA.

TIMSS Data domain and PISA Uncertainty overarching idea

204. It is worthwhile taking a closer look at the Data/Uncertainty content domain, as PISA has considerably more coverage of this content domain (31%) than TIMSS has (14%), and it appears that Western countries have relatively more strengths in this content domain (as compared to the other content domains) than Eastern European and Asian countries have. One may suggest that the Data/Uncertainly domain has a prominence in the PISA test because the need to represent and interpret data is becoming more and more important in everyday lives of the citizens, whether it is reading the newspaper, advertisements, or other forms of communication. Reading graphs and interpreting charts is part of our lives, and not just a school subject. It is therefore hypothesised that skills and knowledge in the Data/Uncertainty domain may be closely related to those in the reading domain, as one area of the reading domain is about document reading. Table 4.10 shows the correlations between country mean scores in TIMSS content domains and PISA reading.

Table 4.10 Correlation between country mean scores in PISA Reading and TIMSS content domains

	Correlation with PISA Reading
TIMSS Number	0.65
TIMSS Algebra	0.62
TIMSS Measurement	0.79
TIMSS Geometry	0.57
TIMSS Data	0.91

205. From Table 4.10, it can be seen that the Data domain stands out as one that is highly correlated with Reading. Since the PISA mathematics test has nearly one third of the items in the Data/Uncertainty domain, it is not surprising that there is a high correlation between PISA mathematics and PISA reading scores. It is also not surprising that country rankings are somewhat difference between TIMSS and PISA, as PISA mathematics, on balance, is testing something a little different from what TIMSS tests.

Examining the spread of PISA and TIMSS achievement distributions

206. The comparisons in the previous sections focus on the differences between country mean scores. While the mean score of a country provides one measure of overall performance of a country, it does not provide a complete picture of country performance. For example, it is often of interest to know how low and high achievers differ within a country and across countries. In this chapter, we examine characteristics of the achievement distributions other than mean scores, such as the spread of the distributions and percentile points.

Standard Deviations

207. Standard deviation is a measure of dispersion (or spread) of a set of data values. The larger the standard deviation, the more spread out the data values are (See Box 4.1 for an explanation). In PISA and TIMSS, the mean and standard deviation of mathematics achievement scores are reported by country. Table 4.11 shows the standard deviations for the 22 countries that participated in both surveys. The last column of Table 4.11 shows PISA standard deviation in TIMSS unit, using the transformation[17] as described earlier in this chapter. The countries are arranged in order of their TIMSS standard deviation, from smallest to largest. If two surveys have the same population definition and similar composition of items, one would expect the standard deviations in the two surveys to have a high correlation. That is, if the mathematics abilities of students in country X are more spread out than those in other countries, then this should be reflected in the standard deviation values in both surveys. A scan down the column of standard deviations in PISA shows that there is not a very strong relationship between TIMSS standard deviation and PISA standard deviation.

208. Figure 4.7 shows a plot of TIMSS standard deviation and PISA standard deviation in TIMSS unit. The correlation between these two variables across the countries is 0.22, a somewhat weak correlation. There are two notable outliers in this graph: Indonesia[18] and Canada–Quebec. The other countries seem to lie somewhere in between these two outliers, with a stronger positive correlation (0.55). The dotted line in Figure 4.7 shows the equality line where TIMSS standard deviation equals PISA standard deviation. It can be seen that 18 out of 22 countries have larger standard deviations in PISA than in TIMSS. A possible explanation for this is that the PISA samples contain students from multiple grades, so there may be a wider spread of mathematics abilities. Of course there are also many other possible reasons for differences in the magnitude of standard deviations between the two surveys. Since the content

17 More specifically, the transformation used was:

PISA standard deviation in TIMSS unit = (PISA standard deviation / 51.92) * 45.08,

where 51.92 was the standard deviation of the PISA country mean scores of the 22 countries participated in both PISA and TIMSS, and 45.08 was the standard deviation of the TIMSS country mean scores.

18 In fact, Indonesia has the highest standard deviation in TIMSS and lowest standard deviation in PISA among the 22 countries. We have sought explanations for this, and it appears that the sampling base (types of schools) may not be the same in the two surveys in Indonesia. A thorough investigation of this is outside the scope of this report. But there are some indications that the PISA and TIMSS samples do not reflect the same population groups in Indonesia, over and above the difference in grade-based and age-based sampling in TIMSS and PISA.

balance and the items are different in the two surveys, one test may spread students out more than the other. However, as PISA has a slightly lower reported reliability than TIMSS' reported reliability, it appears unlikely that PISA test should spread students out more than TIMSS test.

Table 4.11 PISA and TIMSS standard deviations for the 22 countries

	TIMSS standard deviation	PISA standard deviation as reported	PISA standard deviation transformed to TIMSS unit
Quebec, Canada	58	93	81
Tunisia	60	82	71
Basque country, Spain	64	82	72
Ontario, Canada	66	83	72
The Netherlands	69	93	80
Norway	71	92	80
Sweden	71	95	82
Hong Kong - China	72	100	87
Belgium Flemish	73	105	91
Latvia	73	88	76
Scotland	75	84	73
England	77	93	81
Italy	77	96	83
Russian Federation	77	92	80
New Zealand	78	98	85
Hungary	80	94	81
Japan	80	101	87
United States	80	95	83
Australia	82	95	83
Slovak Republic	82	93	81
Korea	84	92	80
Indonesia	89	81	70

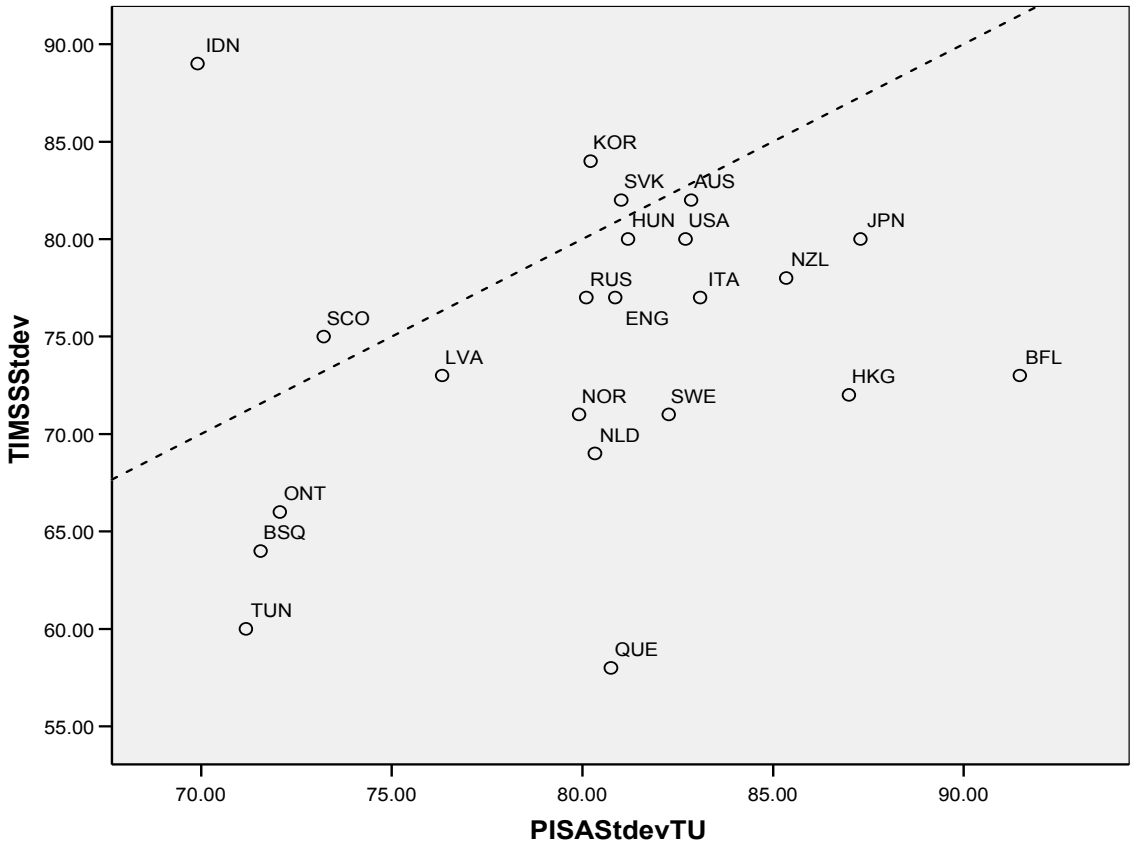


Figure 4.7 TIMSS standard deviation versus PISA standard deviation (in TIMSS unit)

Percentiles

209. To examine the distributions more closely, the 5th and 95th percentile points are shown in Table 4.12. Figure 4.8 shows a country-by-country comparison of PISA and TIMSS percentile points. For each country, two bars are shown. The first bar shows the PISA results, and the second bar shows TIMSS results. The top end of each bar shows the 95th percentile point, and the bottom end of each bar shows the 5th percentile point. The circle in the middle shows country mean score. For example, the pair of bars for Australia shows that the spread of the achievement distribution of Australian students in PISA is similar to the spread of TIMSS achievement distribution, but the whole PISA distribution is shifted upwards as compared to TIMSS distribution. In contrast, for Hong Kong, the spread of the PISA distribution is much larger than the spread of the TIMSS distribution, and the whole PISA ability distribution is shifted downwards as compare to the TIMSS distribution.

Table 4.12 PISA and TIMSS 95th and 5th percentiles for countries participating in both Surveys in 2003

	TIMSS 5th percentile	TIMSS 95th percentile	PISA 5th percentile	PISA 95th percentile	PISA 5th percentile (in TIMSS unit)	PISA 95th percentile (in TIMSS unit)
Australia	368	634	364	676	391	662
Basque country, Spain	379	591	361	631	389	623
Belgium Flemish	398	643	360	707	388	689
England	373	627	354	660	383	648
Hong Kong - China	455	691	374	700	400	682
Hungary	398	656	335	644	366	634
Indonesia	266	558	233	499	278	508
Italy	355	606	307	623	342	616
Japan	433	697	361	690	388	674
Korea	439	715	388	690	412	674
Latvia	386	625	339	626	370	619
New Zealand	364	623	358	682	386	667
Norway	340	573	343	645	373	635
Ontario, Canada	411	628	390	665	413	653
Quebec, Canada	449	640	378	682	403	667
Russian Federation	381	632	319	622	352	615
Scotland	368	615	380	660	405	648
Slovak Republic	371	642	342	648	372	638
Sweden	378	614	353	662	381	650
The Netherlands	417	644	385	683	409	668
Tunisia	316	515	229	501	274	510
United States	369	635	323	638	355	629

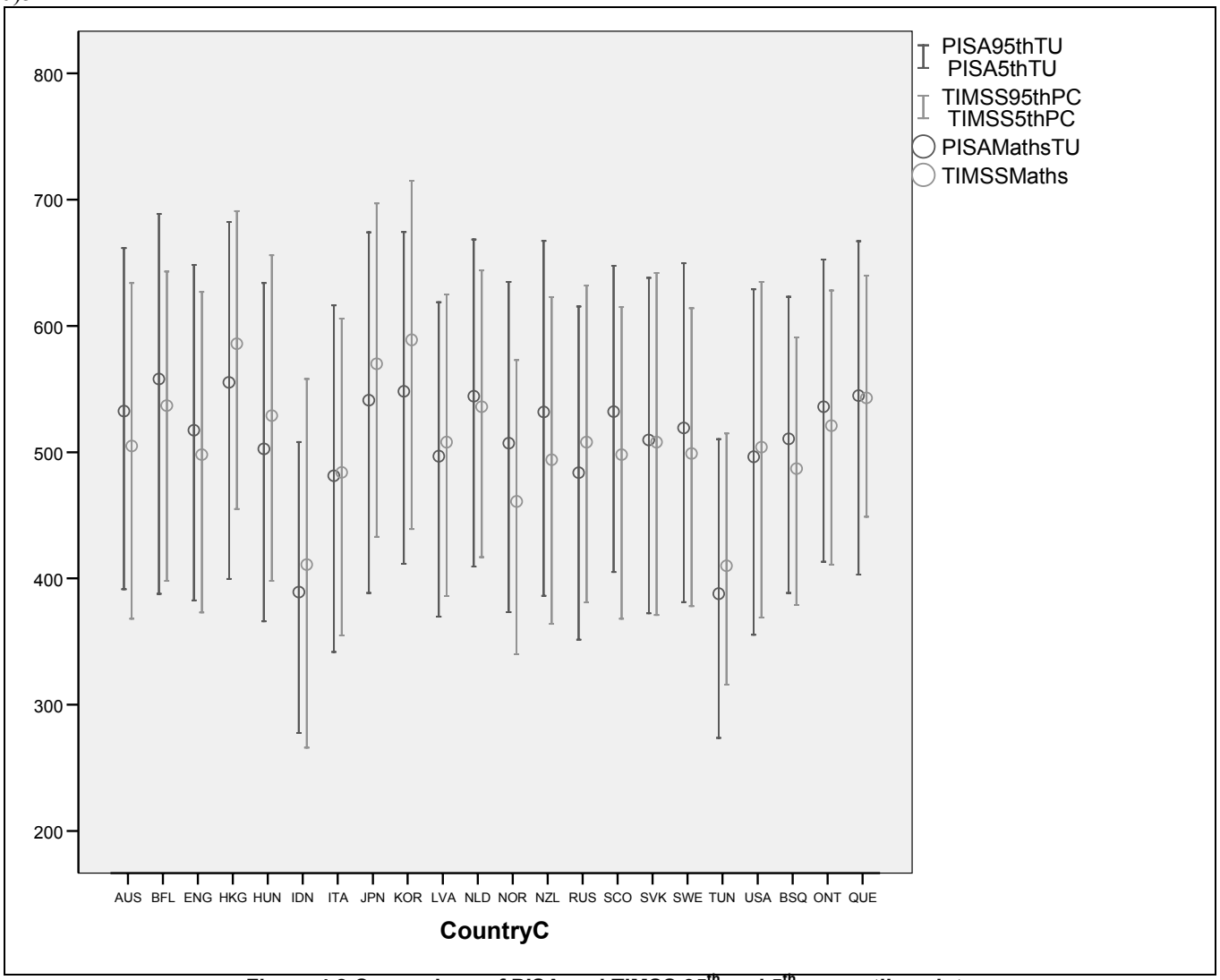


Figure 4.8 Comparison of PISA and TIMSS 95th and 5th percentile points

210. Factors that have an impact on the differential spread of ability distributions in PISA and TIMSS will be difficult to identify, as there are likely between-country differences in terms of characteristics of samples and advantages/disadvantages in content balance. One can only make some hypotheses. For example, in Hong Kong and Tunisia, the spread of the achievement distribution is considerably larger in PISA than in TIMSS. One possible reason for this is that, in both Hong Kong and Tunisia, the PISA sample contains many students from grades lower than the typical grade of 15 year-olds. Figure 4.9 shows the histograms of grade distribution for Hong Kong and Tunisia.

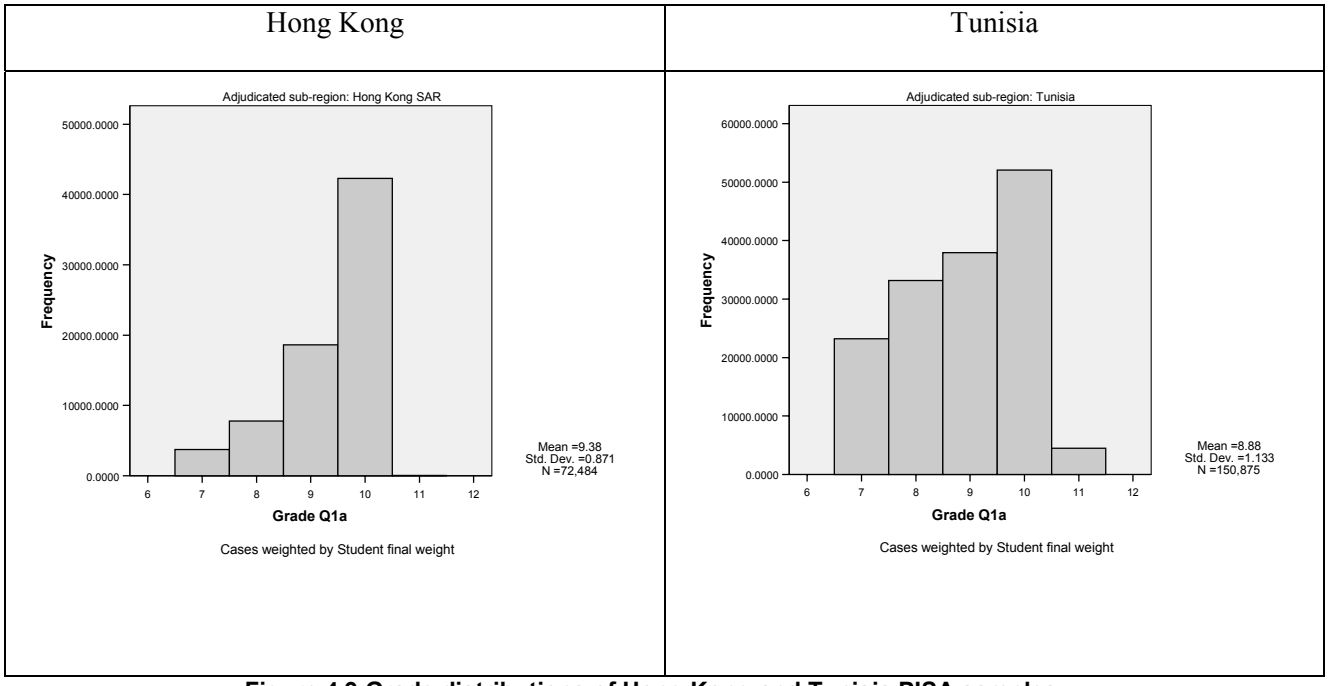


Figure 4.9 Grade distributions of Hong Kong and Tunisia PISA samples

211. The results in this section show that the spread of ability distributions need to be interpreted with regard to the population definition. If one makes a statement that students in country X are more similar in ability than students in country Y, the underlying population definition must be referred to. The comparisons between TIMSS and PISA show that, within a country, the variation in student abilities for a grade level may be quite different from the variation in abilities for an age group. Further, in general, variations of mathematics abilities for an age group tend to be larger than variations for a grade level.

The impact of calculator availability on differences in performance in TIMSS and PISA

212. In Chapter 2, it was found that the percentages of students with access to a calculator differ across countries, and differ between the two surveys. These differences do not have an impact on the differential performance of countries in TIMSS and PISA. The correlation between these two variables is 0.106 (p=0.676). That is, the different policies on calculator use in TIMSS and PISA do not explain the differences in country rankings in TIMSS and PISA.

Summary

213. This chapter explores the factors that may have an impact on the observed differences between country mean scores in PISA and TIMSS. It is found that the age at testing in TIMSS has a positive relationship with the differential performance in TIMSS and PISA. That is, countries with an older cohort of students tend to perform relatively better in TIMSS than in PISA. However, this does not mean that better performance in TIMSS is necessarily due to older students in the TIMSS sample. In fact, there is a

strong negative correlation between *age at time of testing in TIMSS* and *years of schooling at time of testing in PISA*. The older the students are in the TIMSS sample for a country, the fewer number of years of schooling the students have at time of testing in PISA. Therefore, a relatively better performance in PISA could be due to more years of schooling at time of testing in PISA.

214. Further, a Content Advantage Index was constructed based on each country's performance in the five content areas of mathematics in TIMSS, and the relative proportions of items in the content areas in each survey. Given that the content balance of PISA is quite different from that of TIMSS, some countries have more advantage in PISA and some have more advantage in TIMSS, depending on their relative strengths and weaknesses in the content areas. It was shown that the differential performance of countries in PISA and TIMSS was closely related to the content balance of each survey.

215. Using three variables (*TIMSS mathematics country mean score, Age at time of TIMSS testing, Content Advantage Index*) as predictors for PISA mathematics country mean scores, it was found that 93% of the variance of PISA mathematics scores could be explained by these three predictors. Further, as there is more reading demand in the PISA mathematics tests, if PISA reading score is used as an additional explanatory variable, then 97% of the variance of the PISA mathematics scores can be explained. With these four predictor variables, 11 out of the 22 countries have a predicted PISA mathematics score within the confidence interval of the reported PISA score.

216. The standard deviations of achievement distributions in PISA are generally larger than the standard deviations in TIMSS. This is likely the result of different population definitions where PISA samples contain students from multiple grades while TIMSS controls for the grade level.

CHAPTER 5 - GENDER DIFFERENCE AND ATTITUDES

Introduction

217. It will be of interest to compare the results drawn from PISA and TIMSS in terms of the performance of subgroups of students, such as girls and boys. In addition, both PISA and TIMSS collected data on student home background information and attitudes towards mathematics. A comparison of the similarities or differences in the findings will be interesting. If the same results are found, then each study provides some evidence of validity for the other study. If different results are found, it will be interesting to investigate why there are differences, and, in doing so, one hopes to gain a better understanding of the results of each study.

Gender differences

Gender differences for overall mathematics scale

218. Do PISA and TIMSS arrive at the same conclusions about the performance of girls and boys? On the surface, the answer seems to be "No".

219. TIMSS found that

> *On average, across all countries, there was essentially no difference in achievement between boys and girls at either the eighth or fourth grade, although the situation varied from country to country. (IEA, 2003, p.47)*

220. In contrast, in PISA, apart from two countries (Iceland and Thailand) where girls' mean mathematics score is higher than that for boys, in all other 38 countries, boys' mean score is higher, although not all significant statistically. Figure 5.1 shows pictorially the differences between the performance of girls and boys, for TIMSS and PISA, where darkened bars indicate statistical significance (extracted from TIMSS report, p48, (IEA, 2003), and PISA report, p97, (OECD, 2004)).

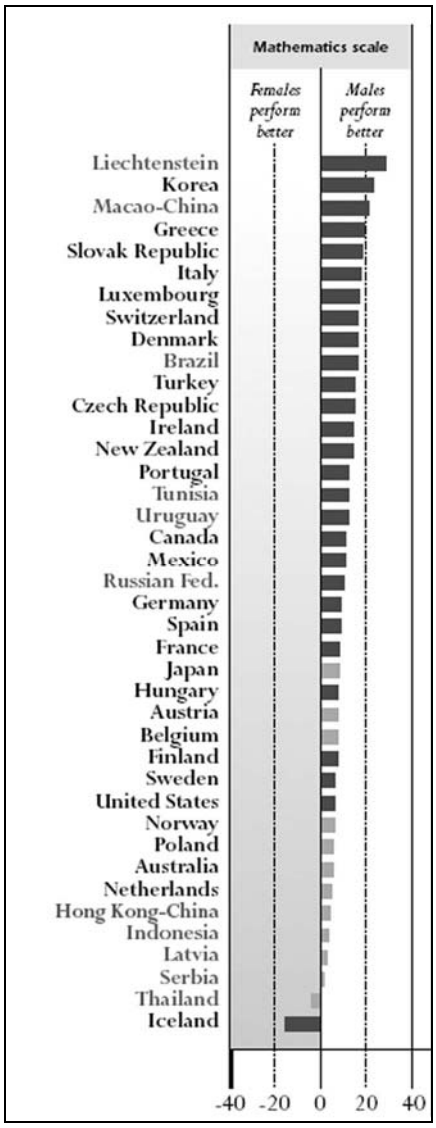
221. On first impression, there appears to be a discrepancy between TIMSS and PISA results of gender differences across countries. In TIMSS, there are about equal numbers of countries where girls performed better and where boys performed better. On the other hand, in PISA, boys performed better in the majority of countries. If one uses TIMSS results, one might conclude that, at Grade eight, there is no clear gender difference in mathematics performance. If one uses PISA results, one might conclude that 15 year-old girls lag behind 15 year-old boys in mathematics performance among OECD countries[19].

19. In 26 out of 40 countries in PISA, boys' performance is statistically significantly better than girls' performance. However, collectively for the OECD population, 38 out of 40 countries showed that the mean score for boys is higher than the mean score for girls. This is extremely significant statistically; (Consider tossing a coin 40 times and obtain a head 38 times. There is little doubt that the coin is biased.)

Figure 5.1 Gender Difference in TIMSS and PISA for overall mathematics scale (Extract from TIMSS and PISA reports)

TIMSS

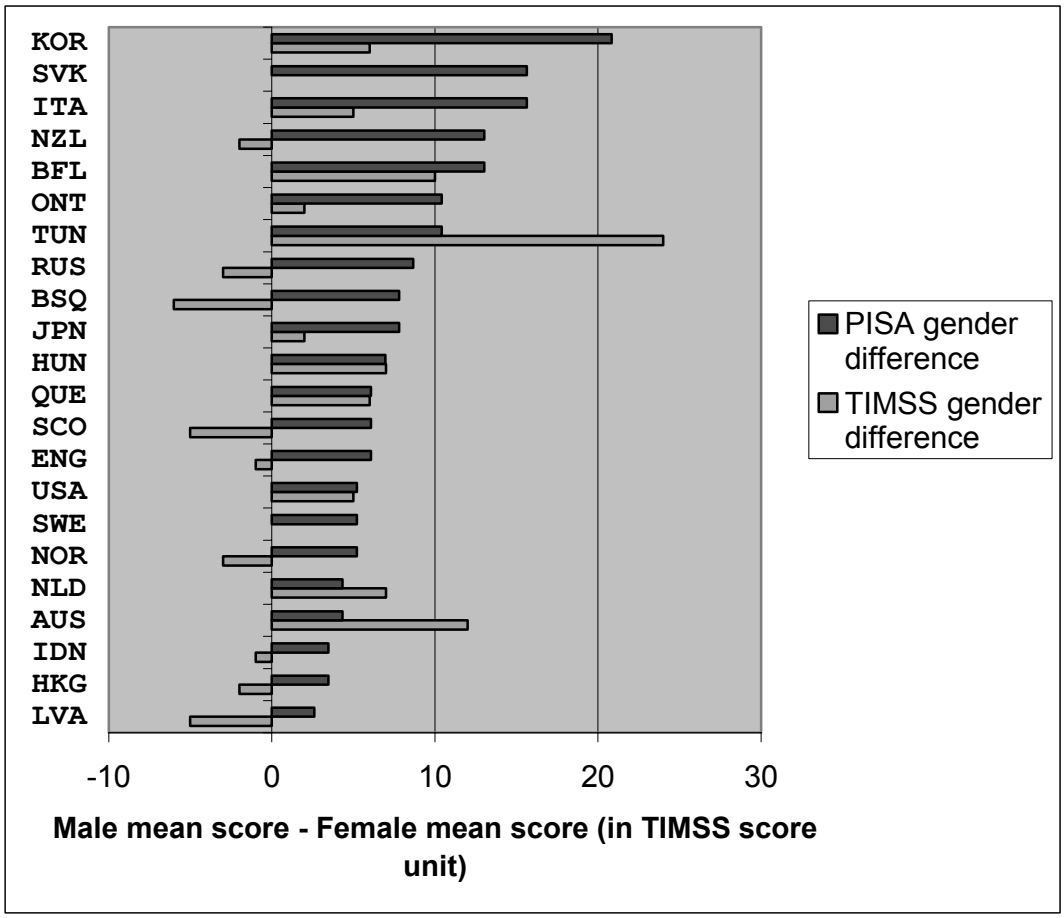
PISA



222. On closer examination, all the countries where girls performed better in TIMSS have not participated in PISA. These countries are Serbia, Macedonia, Armenia, Moldova, Singapore, Philippines, Cyprus, Jordan and Bahrain. None of these countries is an OECD member. On the other hand, in TIMSS, in nine countries boys performed better including the OECD member countries the Flemish Community of Belgium, Hungary, Italy and the United States (the other countries were Chile, Ghana, the Lebanon, Morocco and Tunisia). So it appears that gender difference varies between countries, as noted in both the TIMSS and PISA reports, and that different compositions of participating countries could provide a different picture of overall gender difference in a study. For valid comparisons, one needs to look at the same group of countries.

223. Figure 5.2 shows gender differences for the 22 countries that participated in both PISA and TIMSS.

Figure 5.2 Comparisons of gender differences in PISA and TIMSS



224. In Figure 5.2, the countries are arranged in descending order of the magnitude of gender difference in PISA (in TIMSS score unit). For each country, the first bar shows gender difference (in mean scores) in PISA, and the second bar immediately below the first bar shows gender difference (in mean scores) in TIMSS for that country. If there is a high correlation between PISA and TIMSS gender differences, one would expect the second bar for each country to be longest (and positive) at the top of the graph, and decreases in length or becoming negative, as for the bars for PISA. However, it can be seen that TIMSS gender differences do not diminish as one scans down the graph, as PISA gender differences diminish. The correlation between PISA and TIMSS gender differences is 0.23, indicating that there is a weak relationship between the two variables. Consequently, one cannot conclude that there is a strong agreement between PISA and TIMSS results in terms of gender difference for the overall mathematics scale.

225. It should also be noted that Figure 5.2 shows that, on average, gender difference is larger in PISA than in TIMSS, from a visual comparison of the size and direction of the two bars for each country. That is, on average, boys outperformed girls by a greater amount in PISA than in TIMSS. On average, gender differences in TIMSS are not only smaller in magnitude, but girls outperformed boys in a number of countries.

Gender differences by mathematics content areas

226. A comparison of the composition of PISA and TIMSS tests in terms of the balance of content domains was presented in Chapter 3. It was shown that there was a significant difference in terms of the proportion of items from each traditional content area. How do gender differences vary across mathematics content areas? Both TIMSS and PISA found that gender difference is not uniform across mathematics content areas. In TIMSS, girls performed better in Algebra, while boys performed better in Measurement. In Algebra, girls performed significantly better than boys in 23 countries/regions, while boys performed better in only 3 countries/regions. In measurement, boys performed significantly better than girls in 15 countries/regions, while girls performed better in only 2 countries/regions. For Number, Geometry and Data, the difference in performance between gender groups is not as great as for Algebra and for Measurement. However, there is some suggestion that boys performed a little better than girls did, on average, in these three content areas. Table 5.1 provides a tally of the number of countries/regions in terms of differential performance between boys and girls.

Table 5.1 Comparison of performance of boys and girls in TIMSS by content area

Mathematics content area	Number of countries/regions in which BOYS PERFORMED BETTER	Number of countries/regions in which GIRLS PERFORMED BETTER	Number of countries/regions in which there is NO DIFFERENCE between performance of boys and girls
Number	14	10	26
Algebra	3	23	24
Measurement	15	2	33
Geometry	13	8	29
Data	9	8	33

227. For PISA, gender differences are examined by Overarching Ideas. Table 5.2 shows a summary of the results.

Table 5.2 Comparison of performance of Boys and Girls in PISA by Overarching Ideas

Mathematics content area	Number of countries/regions in which BOYS PERFORMED BETTER	Number of countries/regions in which GIRLS PERFORMED BETTER	Number of countries/regions in which there is NO DIFFERENCE between performance of boys and girls
Space and shape	32	1	7
Change and relationship	21	1	18
Quantity	16	1	23
Uncertainty	29	2	9

228. Although PISA and TIMSS have different proportions of countries in which boys outperformed girls, there is still some consistency in the results from the two studies. If PISA's overarching idea *Space and Shape* can be mapped onto TIMSS' Measurement and Geometry domains (see Table 3.6), then both studies found that boys outperformed girls by the greatest amount in this content area.

229. In PISA, differential performance between boys and girls is least in the area of the overarching idea *Quantity* (although boys still performed significantly better). As PISA's overarching idea *Quantity* can be regarded as a subset of TIMSS' *Number* content domain (see the Quantity section in Chapter 3), it is observed that the gender difference for the content domain *Number* in TIMSS is also less than the difference in *Measurement.*

230. In TIMSS, at the international level, the only content area in which girls outperformed boys is Algebra. In PISA, only 7 items are classified as Algebra. So one might conjecture that most PISA items "favour" boys, since there are not many Algebra items in PISA. This could explain, at least in part, the difference in the findings from the two studies for the overall mathematics scale. That is, in TIMSS, there was little gender difference in overall mathematics performance internationally, while a large difference in PISA was found.

231. In PISA, there is a large gender difference for the Overarching Idea *Uncertainty* (as compared to the other Overarching Ideas), but in TIMSS, the content area *Data* showed less gender difference than, say, for *Number.* Since *Uncertainty* could be regarded as part of the *Data* content area, there seems to be some inconsistencies in the findings here. However, *Data,* in general, covers topics in statistics, while *Uncertainty* deals more with Chance (as in Chance and Data). So it seems that *Uncertainty* in PISA is a narrower domain than *Data* in TIMSS. Some aspects of *Data* in TIMSS may be classified as *Change and Relationships (e.g.,* graphing data) in PISA (see Table 3.6). So one might conclude that, for topics dealing with *Chance,* or, more formally, *Probability,* boys tend to do a great deal better than girls, as found in PISA, but not for *Data.*

232. It should be noted that, while there are considerable variations across countries in gender differences in both PISA and TIMSS, there is a consistent pattern of gender differences in the content areas of mathematics. That is, in general, if one content area shows large gender difference internationally, it is usually reflected in the gender difference within a country, where that content area shows the largest gender difference relative to the other content areas within a country. This suggests that there may be some underlying factors, whether biological or social, that differentiates between boys and girls, so that boys may be naturally better than girls at certain tasks, and vice versa. For example, boys tend to perform better in certain spatial tasks (Voyer, Voyer and Bryden, 1995). Gallagher, Levin and Cahalan (2002) also found that males performed relatively better on items requiring the use of spatial representations, but gender difference is small on items involving the use of mathematics language and recall of knowledge.

233. On the other hand, the vast variation in gender differences across countries indicates that gender difference could be addressed, and there are many countries where gender difference in mathematics performance is negligible. Further studies looking into how some countries address gender equity may be useful. It should be noted that, while, in PISA, significant gender differences were found in mathematics, the magnitude of the difference is much smaller than the gender difference found for Reading (OECD, 2005, p.98). For Reading, girls outperformed boys by a great deal more. From this point of view, it is somewhat surprising that more gender difference was observed in PISA than in TIMSS, since PISA mathematics items require more reading than TIMSS mathematics items, as many TIMSS items are short and context free items. If reading "gets in the way" of successfully responding to a mathematics item, then one would expect the gender difference in PISA mathematics to be less than that observed in TIMSS.

Possible explanations for observed differences between PISA and TIMSS in gender gap

234. In summary, there could be four reasons why the observed gender difference in PISA is larger than in TIMSS:

235. First, in TIMSS and PISA reports, conclusions were drawn from an overall picture of gender difference across countries. Since there are different countries participating in each study, the overall pictures are somewhat different. An observation is that countries where girls' performance is higher than boys' in TIMSS are not OECD member countries. It is possible that OECD countries, being more developed than non-OECD countries in general, may have systematic similarities between themselves

(and differences from non-OECD countries) in education systems that exacerbate the gender difference in mathematics achievement.

236. Second, it has been shown that gender differences are not uniform across mathematics content areas. Since PISA and TIMSS have different content compositions for the tests, gender differences for the overall test are likely to be different for PISA and TIMSS, depending on whether there are more items giving advantage to boys or to girls in a test. From this point of view, gender differences should be interpreted with regard to the composition of a test, and not just for mathematics as a whole.

237. There may also be subtle influences from the choice of particular contexts for individual items which have not been detected by normal vetting procedures.

238. Third, PISA items are more application oriented, while TIMSS items are more curriculum oriented. This difference in the emphasis of the type of the items may have different effects on the performance of girls and boys. Friedman (1989) found that girls are better at curriculum-based items, while boys are better at problem-solving items. That is, girls are not as good as boys in applying mathematical knowledge for functional use in everyday life. In PISA, the correlation between the problem solving domain and the mathematics domain is a little higher for boys than for girls, indicating that there is a slightly more consistent behaviour for boys than for girls in successfully (or unsuccessfully) answering problem solving items and mathematics items in PISA.

239. Fourth, the PISA report described growing gender gap in mathematics achievement as students get older (Box 2.3, p96, OECD, 2004). Since PISA cohort has slightly older students than TIMSS' cohort, it is likely that gender differences for 15 to 16 years are greater than for 14-year-olds, as mathematics becomes a more demanding subject. Friedman (1989) found this result in the meta-analysis and also commented on supporting findings from previous literature. Most commonly with young children (e,g. up to age 10) no differences are found, or if found they favour girls. In the junior high school years, Friedman reported that a mixed pattern develops, with either no differences or differences favouring boys, with the exception that very gifted mathematics students are more likely to be boys. Differences favouring males increase in older age groups. For example, the Australian report of the TIMSS Population 3 advanced mathematics group (Lokan and Greenwood, 2001) noted that in almost all countries and on all topics, there were differences favouring males, often statistically significant, and these differences were much stronger than in the TIMSS studies with younger students.

240. All of the above four points suggest that explanations are worth further investigation to try to understand gender differences in mathematics achievement, and to put in place measures to reduce gender differences within those countries where gender differences are large. Friedman (1989) pointed to a social, rather than a biological, explanation when commenting on the meta-analysis result that gender differences had closed rapidly over two decades in the USA. This means that gender differences can be rectified with appropriate intervention.

Gender difference in the spread of achievement distributions

241. PISA noted that

> *Gender differences tend to be larger at the top end of the performance distribution. (OECD, 2005, p.98)*

242. That is, on average, there are more boys than girls at the top end of the scale. This can be readily seen from Figure 5.3 where the percentages of boys and girls in Level 6 of the mathematics scale are shown. Apart from Iceland where there are more girls than boys in Level 6, in every country where the percentage of students in Level 6 is not zero, there are more boys than girls in Level 6. This finding is

consistent with that of Friedman (1989) who found that often gifted mathematics students are more likely to be boys. In addition, in only 3 of 16 countries in the TIMSS study of advanced mathematics did more females than males do advanced mathematics.

243. For TIMSS, the percentages of students in benchmark levels are not reported separately for boys and girls. However, the standard deviations of performance distributions for girls and boys are reported. These are shown in Figure 5.4.

Figure 5.3 Percentages of boys and girls in Level 6 of PISA mathematics scale

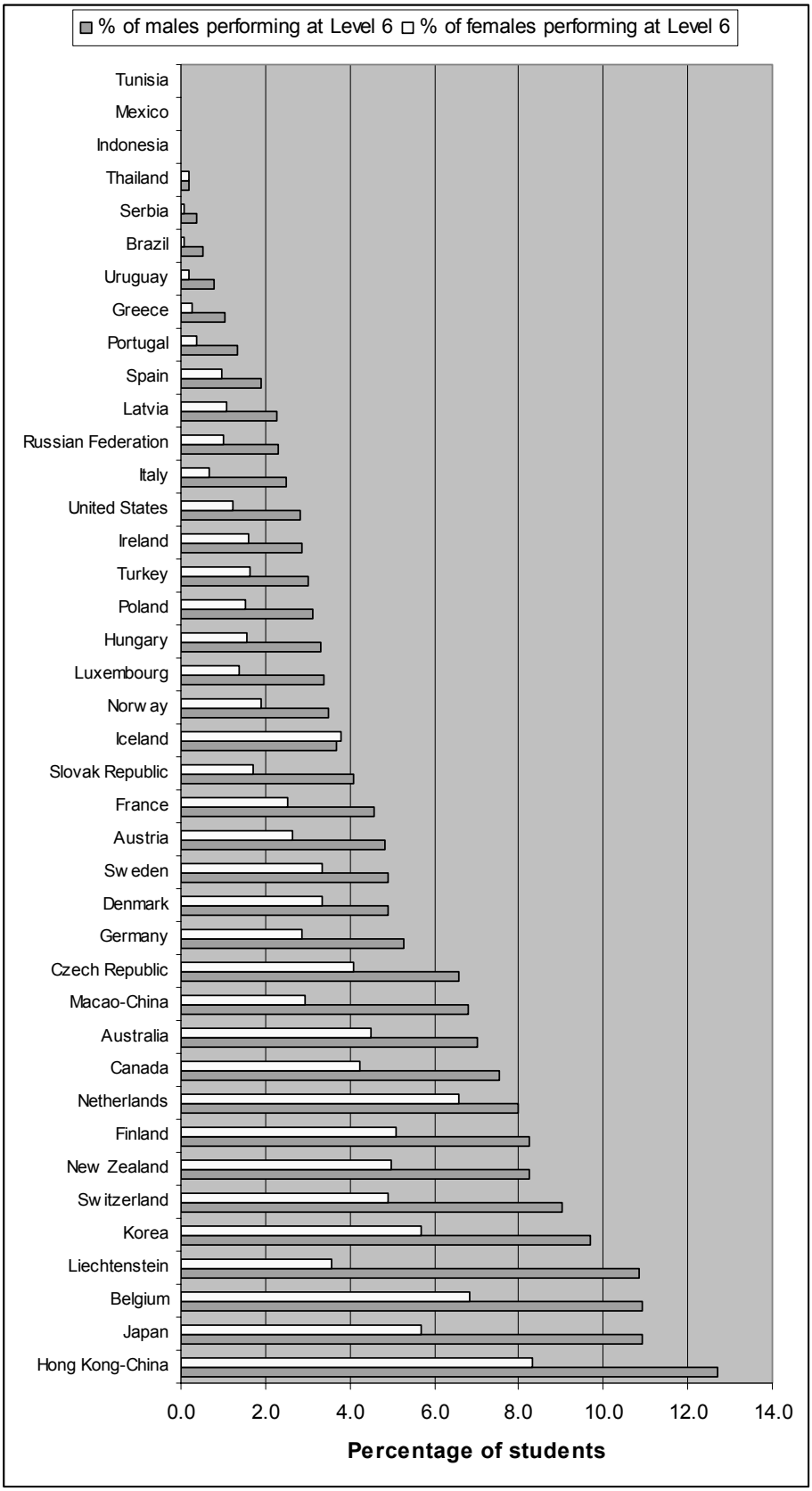
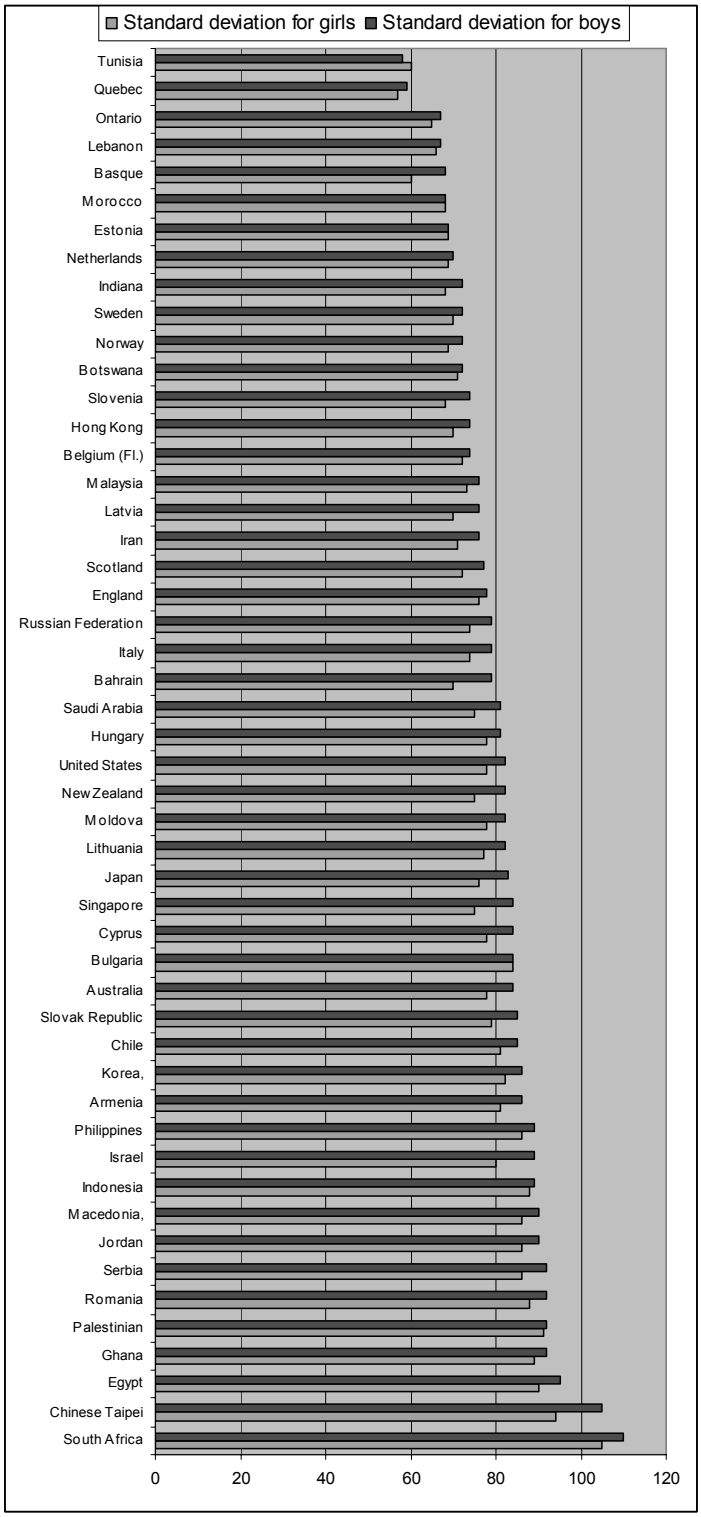


Figure 5.4 TIMSS standard deviations of performance distribution in mathematics for boys and girls



244. The standard deviation of the performance distribution in mathematics for boys is larger than that for girls in 46 out of 50 countries/regions. So it appears that there is some consistency in the findings of PISA and TIMSS in terms of the differences in the spread of performance distributions for boys and girls.

Attitudinal scales

245. Both PISA and TIMSS collect information on students' attitudes towards mathematics. In particular, indices on "self-confidence" and "interest and motivation" have been constructed in both PISA and TIMSS.

Self-confidence index

246. Figure 5.5 shows a comparison of the self-confidence index in TIMSS and the self-concept index in PISA. Note that the vertical axis shows the percentage of students who were assigned to the high level of the TIMSS index. Box 6.1 details how each index was constructed.

247. The correlation between PISA and TIMSS self-confidence indices is 0.73. Indonesia appears an outlier in Figure 5.5. Without Indonesia, the correlation is 0.83. So there is a close relationship between the two indices. Further, in both PISA and TIMSS, it was found that the South East Asian countries (Hong Kong-China, Japan and Korea) had a low self-confidence index, despite the fact that student achievement was high in these countries. On the other hand, English speaking countries tend to have a high self-confidence index in both PISA and TIMSS. Both PISA and TIMSS found that, within each country, students with higher self-confidence index performed better than students with lower self-confidence index.

Box 6.1 Measuring students' confidence in mathematics in PISA and TIMSS

TIMSS: Index of students' self-confidence in learning mathematics (SCM)

This index is based on students' responses to four statements about their mathematics ability:

- I usually do well in mathematics;
- Mathematics is more difficult for me than for many of my classmates*;
- Mathematics is not one of my strengths*;
- I learn quickly in mathematics.

Figure 5.5 presents results for the students with a high level of self-confidence on this index. These students agreed a little or agreed a lot with all four of the above statements on average.

PISA: Index of students' self-concept in mathematics

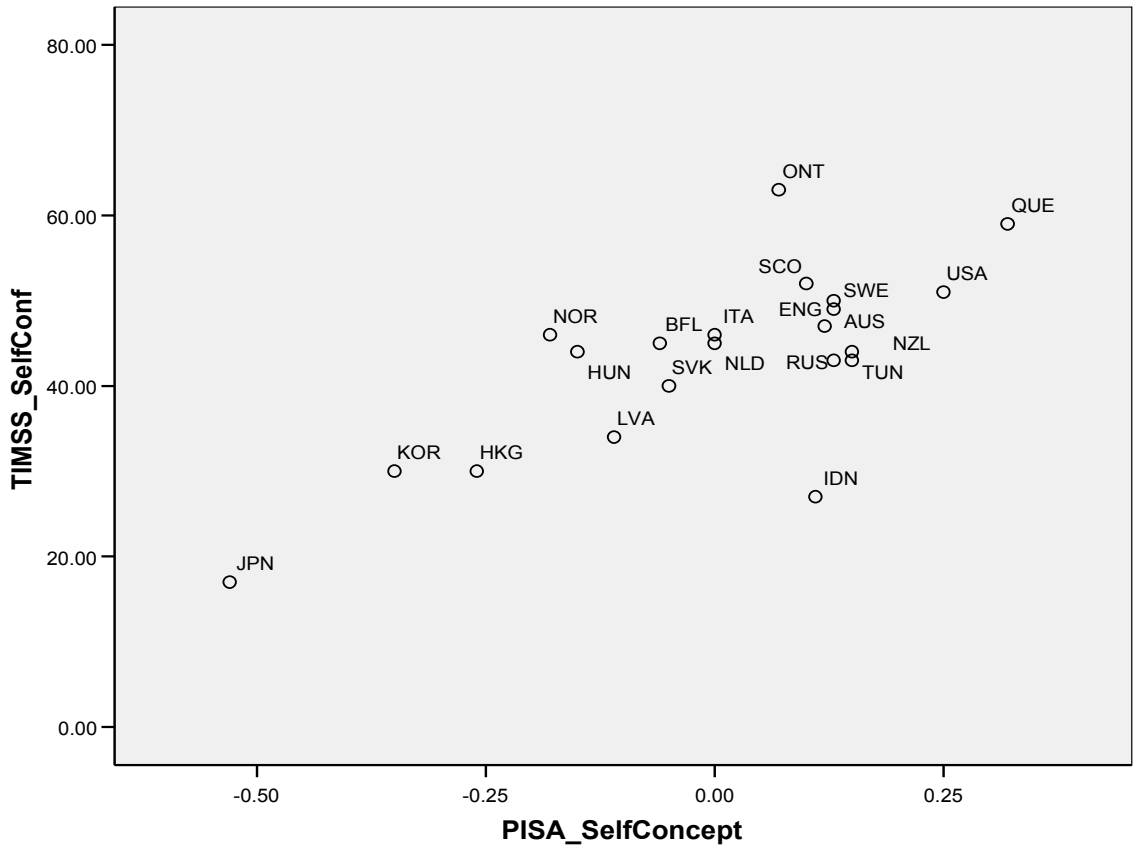
This index is based on students' level of agreement with the following statements about their mathematics ability:

- I am just not good at mathematics*;
- I get good marks in mathematics;
- I learn mathematics quickly;
- I have always believed that mathematics is one of my best subjects;
- In my mathematics class, I understand even the most difficult work.

Students could answer that they strongly agreed, agreed, disagreed or strongly disagreed with the above statements. Positive values on this index indicate a positive self-concept in mathematics.

* The response categories for this statement were reversed in constructing the index.

Figure 5.5 Comparison of PISA and TIMSS self-confidence indices



Interest and motivation indices

248. In TIMSS, an index of "Students' Valuing Mathematics" was constructed from questions about students' interest, enjoyment, and motivation about doing mathematics. In PISA, two separate indices were constructed. The first is "Students' interest in and enjoyment of mathematics". The second is "Students' instrumental motivation in mathematics". For the comparison shown here, PISA's second index is used. Box 6.2 details the components of each index. As can be seen, with the exception of the first two statements in the TIMSS index of valuing mathematics, the two measures are very similar in the way they capture students' motivation to learn mathematics for some external motivating factor whether it be helping them in their current studies, future studies or indeed future work. The first two statements in the TIMSS index of valuing mathematics are similar to components of PISA's index of interest and enjoyment of mathematics, that is, they capture students' enthusiasm for learning the subject itself.

Box 6.2 Measuring students motivation to learn mathematics in PISA and TIMSS

TIMSS: Index of students valuing mathematics

This index is based on students' responses to seven statements about mathematics:

- I would like to take more mathematics in school;
- I enjoy learning mathematics;
- I think learning mathematics will help me in my daily life;
- I need mathematics to learn other school subjects;
- I need to do well in mathematics to get into the university of my choice;
- I would like a job that involved using mathematics;
- I need to do well in mathematics to get the job I want.

Figure 5.6 presents results for the students with a high level on the valuing mathematics index. These students agreed a lot with all seven of the above statements on average.

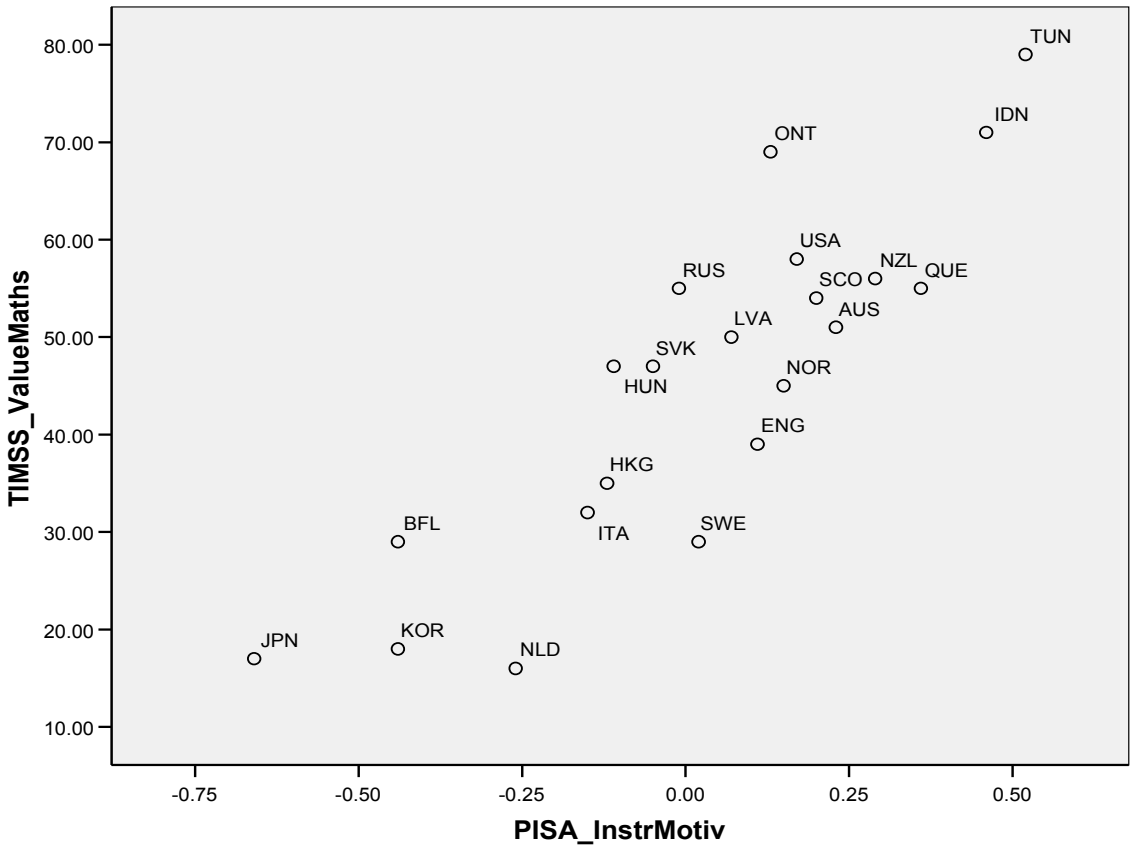
PISA: Index of students' instrumental motivation in mathematics

This index is based on students' level of agreement with the following statements about mathematics:

- Making an effort in mathematics is worth it because it will help me in the work that I want to do later on;
- Learning mathematics is worthwhile for me because it will improve my career prospects;
- Mathematics is an important subject for me because I need it for what I want to study later on ;
- I will learn many things in mathematics that will help me get a job.

Students could answer that they strongly agreed, agreed, disagreed or strongly disagreed with the above statements. Positive values on this index indicate higher levels of instrumental motivation to learn mathematics.

Figure 5.6 Comparison of indices of valuing/motivation in mathematics in TIMSS and PISA



249. Figure 5.6 shows a comparison of PISA and TIMSS indices on valuing/motivation in mathematics. The correlation between TIMSS "Valuing Mathematics index" and PISA "Instrumental Motivation index" is 0.87, showing a close relationship between these two variables.

250. The similarities in the findings of PISA and TIMSS on student motivation and self-confidence in doing mathematics provide some reassurance of the validity of the results. It also suggests that, regardless whether the sample is age based or grade based, and whether the sample consists of 14 or 15-year-olds, profiles of students' attitudes towards mathematics are largely stable.

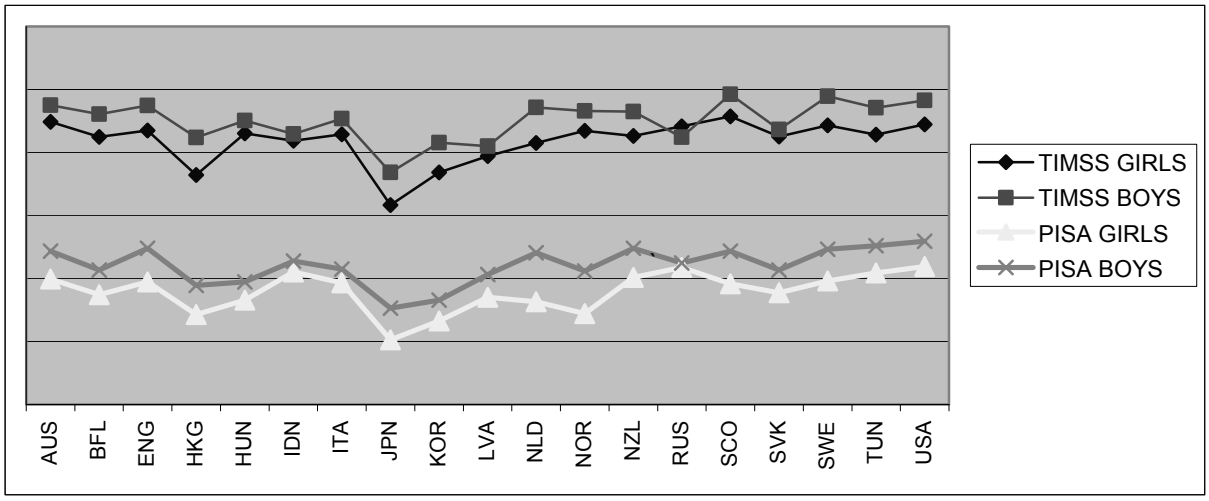
Gender differences in attitudes towards mathematics

251. One of the important findings in PISA is that, despite only moderate differences in mathematics performance between girls and boys, where girls lag a little behind boys on average, there is a large difference between attitudes of girls and boys towards mathematics.

> *A first striking finding is that while gender differences in student performance tend to be modest, there are marked differences between males and females in their interest in and enjoyment of mathematics as well as in their self-related beliefs, emotions and learning strategies related to mathematics. (p.151, OECD, 2004)*

252. PISA made the comparisons between achievement differences and attitude differences based on "effect size"[20] (p.152 – 153, OECD, 2004). In TIMSS, gender differences in attitude are not reported. In this report, the average of TIMSS index of self-confidence was computed separately for girls and boys. These indices are compared to PISA's scale of self-confidence. The results are shown in Figure 5.7[21]. Since the metrics of attitude indices are different in the two studies, the indices have been scaled to have the same variance. So the actual relative positions of the index between PISA and TIMSS is not important (consequently, no vertical scale unit is shown), but the relative size of the gender difference within countries, and the trends across countries, can be compared.

Figure 5.7 Self-confidence in mathematics, by gender



253. A number of observations can be made from the results shown in Figure 5.7. First, the rise and fall of the lines across countries is similar between PISA and TIMSS. This means that countries where students expressed high (or low) self-confidence are consistent in PISA and TIMSS. For example, in Japan and Korea, students rated themselves low on the self-confidence scale in both PISA and TIMSS. This has been discussed in the section on self-confidence index where it is shown that the correlation between the self-confidence index between PISA and TIMSS is high.

254. Second, in both PISA and TIMSS, boys expressed higher self-confidence in mathematics in every country, except for Russia in TIMSS. In PISA, all 19 countries[22] had a higher mean achievement score for boys[23], so it is expected that boys would have higher self-confidence score than girls. However, in TIMSS, girls had a higher mean achievement score[24] in eight of the 19 countries, but girls had lower self-confidence than boys in all countries except for the Russian Federation. This supports the finding in PISA that the gender gap in attitudes towards mathematics is larger than the gender gap in achievement scores.

20. Effect size is a measure of the magnitude of the difference in relation to the variance of all the scores (see Box 3.3, p117, OECD, 2004).

21. Note that data for Spain-Basque country, Canada-Ontario and Canada-Quebec are not available for this graph

22. Data for Canada-Ontario, Canada-Quebec and Spain-Basque country are not available.

23. Although not always statistically significant.

24. Although not always statistically significant.

255. Third, the magnitude of the gender difference in self-confidence for each country is similar between PISA and TIMSS results. For example, in TIMSS, the gender gap in self-confidence is small for Indonesia, Latvia, the Russian Federation and the Slovak Republic. In PISA, these countries also have relatively smaller gender differences. On the other hand, the Netherlands and Sweden have larger gender differences in self-confidence in both PISA and TIMSS. In general, in both studies, Eastern European countries tend to have smaller gender gap in self-confidence, while Western countries and Asian countries have larger gender gap. This is an interesting observation in itself and further investigation will be worthwhile to study how Eastern European countries reduce gender gap in attitudes towards mathematics, or, how Western and Asian countries create gender gap.

256. Fourth, the correlation of the magnitude of gender gap in self-confidence between PISA and TIMSS is moderately high (0.72), as compared to the correlation of the gender gap in mathematics achievement between PISA and TIMSS (0.25). This suggests that attitude scales are more invariant than achievement scales. Achievement scales are more sensitive to variations such as years of schooling and the content balance of the assessments. Importantly this moderately high correlation between the surveys in magnitude of gender gap in self-confidence validates the findings from each survey and suggests that this gender gap is quite established across the ages of 14 to 16 years.

257. Fifth, the correlation between *gender gap in achievement* and *gender gap in self-confidence* across countries is relatively weak in both PISA and TIMSS. That is, in countries where boys outperform girls by a great deal, the difference in self-confidence is not necessarily large. However, in TIMSS, the relationship is slightly stronger than the relationship in PISA, with a correlation of 0.25 for TIMSS, and -0.31 for PISA! This could be due to differences in the nature of the two assessments. While, in TIMSS, the test is more curriculum-based, so performance on the test has a closer relationship with teaching and learning in schools, which in turn relates to students' self-confidence as they work with mathematics in schools, while, in PISA, the test is not so curricular focused. The extent to which students can solve PISA items may not be as closely related to school-based instructions as in TIMSS.

258. In this section, only the self-confidence index is examined. It is expected, however, that other attitudinal indices will provide similar findings.

Socio-economic Background of Students

259. Both PISA and TIMSS collected information on a number of student background variables. PISA reported students' performance in relation to parental occupation, parental education, possessions, single parent family, immigrant status and language spoken at home. TIMSS reported student performance in relation to parental education, language spoken at home, number of books in the home, availability of study desk/table in the home, and the use of computers at home and at other places.

260. In relation to parental education, TIMSS found that, on average, for the Grade 8 cohort, students with university-educated parents[25] scored more than 90 points higher than students whose parents had primary or lower education (IEA, 2003, p.127). In PISA, on average across OECD countries, students whose fathers completed tertiary education scored around 90 points higher than students whose fathers completed primary or lower secondary education (OECD, 2004, Table 4.2c). While the score units are not exactly the same in PISA and TIMSS, and the composition of the countries are different in the two surveys, it is still evident that both surveys found a significant impact of parental education on student performance in mathematics, and the magnitude of the impact is also of similar order of magnitude.

25. This is defined as the highest education level of either parent

261. In relation to home possessions, TIMSS reported student performance against each type of home possessions (the number of books at home, the provision of a desk/table, and the availability of a computer at home) separately. In contrast, PISA constructed an index of cultural possessions from the number of classical literature, books of poetry and works of art in students' homes (OECD, 2004, p.166; OECD, 2004, Table 4.2d; OECD, 2005, p.283). A direct comparison of results is difficult because different variables were used in PISA and TIMSS. However, both PISA and TIMSS found a strong relationship between home possession and student performance.

262. The PISA report (OECD, 2004) provided a more detailed analysis on the impact of socio-economic status on students' performance, both at individual student level and at the school level.

263. Despite differences in the collection of student background variables, the message is clear from both PISA and TIMSS that student background has a significant impact on achievement level.

Summary

264. This chapter explores similarities and differences in the findings of PISA and TIMSS in relation to student backgrounds and attitudes towards mathematics.

265. On the surface, it appears the PISA found that boys performed better than girls, while TIMSS found little gender differences for the grade 8 cohort. On a closer examination of the results, however, the different findings in PISA and TIMSS can be attributed to different cohort of countries and different content balance in the respective tests. In particular, girls performed better in algebra in TIMSS, but there were very few algebra items in PISA. The pattern of gender differences is consistent between PISA and TIMSS findings. For example, both TIMSS and PISA found that boys outperformed girls by the greatest amount in the content areas of measurement and geometry (or space and shape as labeled in PISA). In contrast, both TIMSS and PISA found the smallest gender difference for the number content area (or quantity as labeled in PISA).

266. Further, in both TIMSS and PISA, the spread of mathematics achievement distribution is larger for boys than for girls.

267. Both PISA and TIMSS constructed indices of students' self-confidence. There is a strong agreement in the measures of this index at the country level between PISA and TIMSS. Similarly, TIMSS' Valuing index and PISA's Motivation index have a correlation of 0.89 between country means, showing a close relationship between these two variables. In short, there is a good agreement between PISA and TIMSS on the measures of students' attitudes towards mathematics, despite different age groups in the two surveys.

268. Information on students' socio-economic background was also collected in both PISA and TIMSS. Both surveys found that socio-economic background had a significant impact on students' performance in mathematics.

CHAPTER 6 - CONCLUSIONS

Overview

269. The concurrent testing of TIMSS and PISA in 2003 provided education researchers with a unique opportunity to gain an in-depth understanding of the results of international large-scale assessments. Comparisons of the methodologies and the results between the two surveys led to a number of important findings that not only would help us interpret the survey results, but also would help us plan future surveys. This chapter discusses the findings and their implications for current and future PISA and TIMSS surveys.

270. The most important finding in this report is that differential performances of countries in TIMSS and PISA can be accounted for by a number of factors. In particular, *test content balance* is the most significant factor. Other factors include *years of schooling* and *reading load in items*. The identification of factors not only enables us to draw valid conclusions from country scores and rankings, it also provides us with evidence that both PISA and TIMSS conducted the surveys with rigorous procedures, since the results can be cross verified and validated across the two surveys. Without two concurrent surveys, it would not be possible to examine reliability and validity to this extent. Some specific findings and their implications are discussed below.

The impact of content balance on achievement results

271. A re-classification of PISA items according to TIMSS content domains (*number*, *algebra*, *measurement*, *geometry* and *data*) shows that PISA and TIMSS have quite different test content balance. PISA has more *number* and *data* items, and fewer *algebra* items. The differences in test content balance between the two surveys account for 66% of the variation of country mean score differences between PISA and TIMSS. This finding suggests that mathematics curriculum coverage has a significant impact on student performance in PISA and TIMSS, and it is consistent with TIMSS' finding that different countries have different strengths and weaknesses in the five mathematics content areas (Exhibit 3.1, IEA, 2003).

272. A link between students' performance and a country's curriculum is a reasonable conjecture, but it is not readily established, as it involves an extensive survey of the curriculum of each country. TIMSS conducted a survey of the mathematics curriculum in each country. However, the results of the curriculum survey were not examined closely in relation to the relative performance of each country in each mathematics topic. There is only one note in the TIMSS 2003 International Mathematics Report (IEA, 2003) about the link between curriculum coverage and achievement results:

> Although the relationship between inclusion in the intended curriculum and student achievement was not perfect, it was notable that several of the higher-performing countries reported high levels of emphasis on the mathematics topics in their intended curricula and that those with the lowest levels of curricular coverage came from the lower half of the achievement distribution. (p.182, IEA, 2003)

273. Below we take a look at the relationship between the implemented curriculum (as represented by instructional time) and achievement scores.

274. A survey of the percentage of instructional time in mathematics class devoted to TIMSS content areas shows some variations across countries (Exhibit 7.4, IEA, 2003). For example, in the Russian

Federation, only 3% of time in mathematics class is devoted to the *data* strand at Grade 8 level, while in Australia, 14% of time is devoted to the *data* strand. One might hypothesise that such differences in curriculum emphases of different mathematics content areas will lead to differential student performances in the content areas. To test this hypothesis, the correlation between *Percentage time in mathematics class devoted to a content strand* and *Deviation of achievement score in TIMSS content strand from TIMSS mean score* is computed and presented in Table 6.1.

Table 6.1 Correlation between percentage of instructional time and achievement score in TIMSS 2003 expressed as deviation from the TIMSS mean score over 20[26] countries

	Correlation	p
Number	-0.043	0.87
Algebra	0.540	0.045
Measurement	0.223	0.389
Geometry	0.708	0.001

275. The figures in Table 6.1 show that, in *geometry*, instructional time has a high correlation with the TIMSS achievement score. In *algebra* and *data*, the relationship between instructional time and achievement score is moderate. In *measurement*, the relationship is low. In *number*, there appears to be no relationship between the amount of instructional time and achievement score! These results may be consistent with the view that *number*, *measurement* and *data* are topics which one encounters frequently in everyday life, but *algebra* and *geometry* are topics that need to be studied formally to understand the concepts and the use of special terminology, symbols and formulae. In PISA, there are few *algebra* and *geometry* items. Consequently, it could be said that the PISA test contains items that require less direct instruction on the specific mathematics content. From this point of view, the PISA achievement score reflects everyday use of mathematics, which may or may not be learned at schools, while TIMSS achievement score reflects more school mathematics. It should also be noted that instructional time in Table 6.1 is for Grade 8. Some countries teach topics such as *number* and *measurement* mostly in primary schools, so that instructional time at Grade 8 level could be quite little. But this does not necessarily indicate that the topics are not taught. Rather, they are taught in earlier grades.

276. Accordingly, a country with a high score in PISA shows that the students are good at "everyday mathematics", while a high score in TIMSS shows that the students are good at "school mathematics". For example, in terms of country mean scores, Australia performed significantly better than Hungary in PISA, but Hungary performed significantly better than Australia in TIMSS. One might conclude that students in Hungary are better at typical school mathematics than Australian students, but they are not as good as Australian students in using mathematics for everyday life applications. The fact that there are differences in country rankings between PISA and TIMSS results suggests that, at least in some countries, school mathematics has not prepared students as well in the application of mathematics as in academic mathematics. Conversely, there are countries that have not prepared students as well in specialist areas of mathematics, such as *algebra* and *geometry,* as they have prepared students in solving mathematics problems in everyday life. The question as to which approach is better or which curriculum balance is the best will be for the education policy makers in each country to consider in their own context, and, certainly, neither PISA nor TIMSS alone should set the directions for future mathematics curriculum reform.

277. As content balance has such a significant impact on country performances in mathematics, to establish reliable trend indicators from one survey cycle to another, each survey will need to maintain a stable frame of reference in terms of content balance. This may be more difficult to achieve in PISA than

26 Note that there are no data for the instructional time by content domain for England and Scotland.

in TIMSS, since PISA does not examine content balance in terms of traditional mathematics curriculum strands. Given that there is not a one-to-one mapping between the Overarching Ideas and traditional mathematics curriculum strands, the PISA content balance in terms of traditional mathematics curriculum strands may vary across different cycles of PISA. This issue needs to be monitored carefully. On the other hand, the content balance of an assessment cannot be static over many years, since the world will change over time. What is relevant now may no longer be relevant in ten years' time. So test content must change accordingly. One can already see changes in TIMSS, such as the permission to use calculators in the tests in 2003.

278. Even if the test contents can stay relatively stable across survey cycles, the curricula of countries change continually (Leung, Graf & Lopez-real, 2006). TIMSS (IEA, 2003, p.165) also reported that the curricula in participating countries were being revised constantly. This in turn will affect trend estimates for the countries. Therefore, as a note of caution, in assessing trends across survey cycles, factors such as content changes in the tests or curriculum changes within countries must be taken into account.

279. The impact of content balance on achievement also casts some doubts about the reliable measure of growth between two grade levels, particularly if the grade levels are far apart, since the alignment of content between different grade levels will be difficult, as topics and content inevitably change across school grades. For example, algebra is usually not taught until secondary schools. Arithmetic computations are more the focus of primary schools but not secondary schools. In primary schools, students mostly work with concrete representations while secondary school students often work with abstract representations. These differences in content balance between primary and secondary school mathematics will make any measure of growth difficult when students cannot easily make use of their knowledge and skills across different content areas.

280. More generally, one message for developers of assessments of mathematics is that the issue of content balance should receive careful consideration and discussion, as performance results may be sensitive to the composition of items from different content areas. The assumption made in many assessment programmes about the *uni-dimensionality* of mathematics items may need careful checking. That is, there is some evidence that mathematics items in PISA and TIMSS measure multiple abilities than a single ability.

The impact of reading load on mathematics achievement results

281. Leaving aside the three South-East Asian countries (Japan, Korea and Hong Kong-China), PISA reading country mean score is found to be a good predictor of the differences between TIMSS and PISA scores. As there is a considerable amount of reading in PISA tests compared with TIMSS tests, it is reasonable to assume that poor readers will not perform as well in PISA as in TIMSS. It is also possible that, in countries where reading achievement is relatively higher, students may be exposed to an environment which facilitates mathematics problem-solving skills in everyday life. For example, students may read newspapers more often, and be familiar with presentations of charts and diagrams for summary of information, or, students may have more opportunity to be intelligent consumers such as through reading advertisements of mobile phone cost plans. Unfortunately, it is difficult to disentangle these factors from the current datasets.

Grade-based and age-based samples

282. It has been contentious whether a grade-based sample or an age-based sample provides the most comparable results in international surveys. The proponents for an age-based sample (OECD, 2004, p.27) argue that, as each country has different school systems with different number of years of pre-schools, it is difficult to define a grade at which the number of years of schooling is aligned across countries. In

contrast, an age-based sample defines the population according to an age range, where age is always clearly defined. However, in the case of an age-based sample, the number of years of schooling will clearly be different across countries. Proponents for a grade-based sample (IEA, 2003, p18) argue that there is a better chance to align the number of years of schooling with grade-based sample. Further, a grade-based sample provides the opportunity to collect information about classroom instructions and implemented curriculum for comparison across countries.

283. This report shows that the age at time of testing of TIMSS can be matched with the number of years of schooling in PISA for most countries. This finding shows that both the TIMSS sample and the PISA sample provide the same degree of comparability across countries, in that neither provides a more stable frame of reference than the other, since both samples can be cross-checked in age and grade.

284. However, the number of years of schooling does have an impact on mathematics achievement. One year of schooling could increase the average achievement by 20 to 40 score points. Therefore, in comparing the differences between TIMSS and PISA results, the number of years of schooling in PISA should be taken into account. A recommendation from this report is that both PISA and TIMSS should endeavour to obtain a better measure of the number of years of schooling, as even a fraction of a year of schooling will make some differences to the achievement scores. Since the rough measure of the number of years of schooling constructed in this report already can provide some analytic power in explaining achievement differences in TIMSS and PISA, any refinement of the rough estimates produced in this report should be an improvement.

Correlated factors

285. While we have identified a number of factors that have an impact on achievement scores, the picture is not so clear as to the causal relationships between achievement levels and the identified factors. It appears that many factors are correlated, sometimes for no clear and obvious reasons. For example, it appears that students in Asian and Eastern European countries tend to start school later than Western countries. These same countries also tend to have a curriculum that is oriented towards an emphasis on formal mathematics. A check on the correlation between the variable *TIMSS advantage index* and *TIMSS age at time of testing* shows a correlation significantly different from zero (correlation=0.487, p=0.035). It is possible that cultural factors could have an impact on both the school systems (in terms of age of entry, etc.) and the curriculum (in terms of formal mathematics versus mathematics for application, etc.), so that groups of countries have similar profiles in a number of aspects of mathematics education.

Gender and attitudinal differences

286. The comparisons of gender differences in PISA and TIMSS in this report (Chapter 5) highlight at least two important points. First, gender differences are not uniform across mathematics content domains. In particular, for *spatial* tasks and *measurement* tasks, boys outperform girls by the greatest amount. This finding has implications for instructional measures to address gender differences. For example, particular kinds of tasks may be designed to engage girls and boys in different ways. Second, gender differences are not uniform across countries. Is there a hint of suggestion that boys outperform girls in mathematics in more developed countries such as OECD member countries? If so, this would seem contrary to expectations that more developed countries would generally have better addressed the equity issue between girls and boys. Is there a stereo-typing promoted in developed countries? The contrast in TIMSS and PISA results regarding gender differences prompts us to examine gender issues in further depths.

287. Both PISA and TIMSS reported similar findings in students' attitudes towards mathematics. The most important one is that there is a larger gender gap in attitudes towards mathematics than in

mathematics achievement, with boys feeling more positive and confident towards mathematics than girls. The combined results of TIMSS and PISA provide us with a more complete picture of students' attitudes towards mathematics. In particular, there is still much to learn by Western countries from Eastern European countries about reducing the gender gap in students' attitudes towards mathematics.

And finally…

288. A number of important findings in this report could only be made when there are two international surveys conducted at the same time. From this point of view, the two surveys contribute not only to cross-national comparisons, but also to the validation of similarities and the identification of differences in the results of the two surveys. The interpretations of the results from each survey are enhanced by the cross validation by the other survey. It is comforting to all involved in the surveys to know that the results between the two surveys can be quantitatively validated and crosschecked. It shows that the surveys are conducted with rigour and sound theoretical underpinnings.

289. However, the fact that the results from the two surveys can be corroborated quantitatively suggests that there is some duplication in the outcomes of the surveys, at least in the results in the international reports. The findings in this report highlight the importance of making careful interpretations of results for each country individually, since there are country specific factors that impact on student performance. Consequently, international reports are just the first steps in presenting data collected in these surveys. National reports should be viewed with greater importance for examining the data in more depth in relation to cultural and policy factors specific to each country.

290. In making comparisons between PISA and TIMSS results at the international level, it should be remembered that PISA results reflect relative performance of countries on a set of desirable mathematical proficiencies for everyday life (as set through a consensus building process guided by the PISA mathematics expert group). In contrast, TIMSS results reflect how well countries performed against important areas of current mathematics curricula as agreed by participating countries. The fact that these two approaches to building the mathematics frameworks do not produce the same assessment results is food for thought for both curriculum developers and assessment practitioners.

291. Looking ahead, for future cycles of PISA and TIMSS, many issues raised in this report should be considered in order to provide the best strategies in producing internationally useful and valid results with maximum efficiency for the countries. This calls for collaboration between the two surveys so that the results from the surveys can complement each other, rather than duplicate each other.

References

Adams, R. J. (2003). Response to 'Cautions on OECD's recent educational survey (PISA)', *Oxford Review of Education, Vol. 29, No. 3, September 2003*. pp.377-398.

Bonotto, C. (2003). Suspension of sense-making in mathematical word problem solving: A possible remedy. Taken from http://math.unipa.it/~grim/Jbonotto on 16/8/2003

Brown, G., Micklewright, J., Schnepf, S., & Waldmann, R. (2005). *Cross-national surveys of learning achievement: How robust are the findings?* Retrieved December 28, 2005 from ftp://repec.iza.org/RePEc/Discussionpaper/dp1652.pdf.

Committee of Inquiry into the Teaching of Mathematics in Schools (1982). *Mathematics Counts.* (the Cockcroft report), Her Majesty's Stationery Office, London, United Kingdom.

De Lange, J. (1996). Using and applying mathematics in education. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 49-98). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum Associates.

Ferrini-Mundy, J. (2002). *Draft commentary on PISA 2003 mathematics framework.* Internal report commissioned by OECD.

Freudenthal, H. (1973). *Mathematics as Educational Task,* D. Reidel, Dordrecht, Netherlands.

Friedman, Lynn (1989) Mathematics and the gender gap: A meta-analysis of recent studies on sex difference in mathematical tasks. Review of Educational research, Summer 1989, 59(2), 185 – 213

Gallagher, A., Levin, J., & Cahalan, C. (2002). *Cognitive patterns of gender differences on mathematics admissions tests.* GRE Research, September 2002, GRE Board Professional Report no. 96-17P, ETS Research Report 02-19. Princeton, NJ: ETS.

Gee, J. (1998). *Preamble to a literacy program.* Department of Curriculum and Instruction. Madison, WI.

Gravemeijer, K. (1999). How emergent models may foster the constitution of formal mathematics. *Mathematical Thinking and Learning. An International Journal, 1(2),* 155-177.

Grünbaum, B. (1985). Geometry Strikes Again", *Mathematics Magazine*, 58 (1) , pp 12-18.

IEA (2003). *TIMSS 2003 International Mathematics Report.* Chestnut Hill, M.A: TIMSS International Study Centre.

IEA (2003b). *TIMSS Assessment Frameworks and Specifications 2003.* Chestnut Hill, M.A: TIMSS International Study Centre.

IEA (2004). *TIMSS 2003 Technical Report.* Chestnut Hill, M.A: TIMSS International Study Centre.

Leung, F.K.S., Graf, K-D., & Lopez-Real, F.J. (Eds.) (2006). *Mathematics Education in Different Cultural Traditions – A Comparative Study of East Asia and the West.* The 13th ICMI Study. Springer.

LOGSE (1990). *Ley de Ordenacion General del Sistema Educativo,* Madrid, Spain.

Lokan, J. & Greenwood, L. (2001) TIMSS Maths & Science on the line. Australian Year 12 students' performance in the Third International Mathematics and Science Study. ACER, Melbourne.

Mathematical Sciences Education Board (MSEB) (1990). *Reshaping School Mathematics: A Philosophy and Framework of Curriculum*. National Academy Press, Washington, DC.

Mislevy, R.J. (1991). Randomization-based inference about latent variable from complex samples. *Psychometrika, 56,* Psychometric Society, Greensboro, pp. 177-196.

Mislevy, R.J., & Sheehan, K.M. (1987). Marginal estimation procedures, in A.E. Beaton (Ed.), *The NAEP 1983-1984 Technical Report (Report No. 15-TR-20*), Educational Testing Service, Princeton, N.J.

Mullis, I., Martin, M., Smith, T. A., Garden, R. A., Gregory, K. D., Gonzales, E. J., Chrostowski, S. J., & O'Connor, K. M. (2001). *TIMSS assessment frameworks and specifications 2003.* Chestnut Hill: ISC, Boston College.

Nagasaki, E., & Senuma, H. (2002). TIMSS mathematics results: A Japanese perspective. In D. F. Robitaille & A. E. Beaton (Eds.), *Secondary analysis of the TIMSS data*. (pp.81-93). Dordrecht, The Netherlands: Kluwer Academic Publishers.

National Council of Teachers of Mathematics (NCTM) (1989). *Curriculum and Evaluation Standards for School Mathematics. NCTM*, Reston, VA.

National Council of Teachers of Mathematics (NCTM) (2000). *Principles and Standards for Mathematics.* NCTM, Reston, VA.

OECD (2001). *Knowledge and skills for life. First results from PISA 2000.* Paris: OECD.

OECD (2003). *The PISA 2003 assessment framework.* Paris: OECD.

OECD (2004). *Learning for Tomorrow's World.* Paris: OECD.

OECD (2005). *PISA 2003 Technical Report.* Paris: OECD.

Prais, S. J. (2003). Cautions on OECD's recent educational survey (PISA). *Oxford Review of Education, Vol. 29, No. 2, June 2003.*

Robitaille, D. F., et al (1993*). TIMSS Monograph No. 1: Curriculum frameworks for mathematics and science.* Vancouver, BC: Pacific Educational Press.

Romberg, T., & de Lange, J. (1998). *Mathematics in context.* Chicago: Britannica Mathematics System.

Routitsky, A., & Turner, R. (2003). *Item format types and their influence on cross-national comparisons of student performance,* Paper presented at the AERA Annual Meeting, Chicago, April, 2003.

Routitsky, A., & Zammit, S. (2001). What we can learn from TIMSS: Comparison of Australian and Russian TIMSS-R results in algebra. In H. Chick, K. Stacey, Jill Vincent, & John Vincent. (Eds.), *Proceedings of the 12th ICMI study conference: The Future of the teaching and learning of algebra.* Melbourne, Australia: University of Melbourne.

Routitsky, A., Zammit, S. A (2002) Association between intended and attained Algebra curriculum in TIMSS 1998/1999 for ten countries. *Proceedings of the 2002 annual conference of the Australian Association for Research in Education, Brisbaine.* Retrieved January 2003 from: http://www.aare.edu.au/indexpap.htm (rou02147)

Silver, E. A. (2002). *Review of the proposed OECD/PISA 2003 mathematics framework.* Internal report commissioned by OECD.

van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory.* New York: Springer-Verlag.

Verschaffel, L., Greer, B. & de Corte E. (2000). Making sense of word problems. Swets & Zeitlinger, Lisse.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. Psychological Bulletin, 117(2), 250-270.

Wang, W C. (2000). *PISA Maths analysis report.* Unpublished Internal Report. Camberwell: ACER.

Wu, M. L. (2005). The Impact of PISA in Mathematics Education –Linking Mathematics and the Real World. *Education Journal (Special issue: Analyzing the quality of education in Hong Kong from an international perspective), Volume 32, Number 1, Summer 2004, pp121-140.*

Wu, M. L. (2005b). The role of plausible values in large-scale surveys. Postlethwaite (Ed.). *Special Issue of Studies in Educational Evaluation (SEE) in memory of R M Wolf. 31 (2005)* 114-128.

Wu, M. L. (2006). *A comparison of mathematics performance between East and West – What PISA and TIMSS can tell us.* ICMI Study 13**, Springer, 239-259.**

Zabulionis, A. (2001). Similarity of Mathematics and Science Achievement of Various Nations. *Education Policy Analysis Archives*, *9*(33). Retrieved 15 March 2003 from the World Wide Web: http://epaa.asu.edu/epaa/v9n33/.