行政院國家科學委員會專題研究計畫 成果報告

學生在TIMSS不同數學題型上的成就及其與性別、數學學習情意、學習情境的交互作用(2/2)

研究成果報告(完整版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 95-2522-S-004-001-
執 行 期 間 ：95年08月01日至96年09月30日
執 行 單 位 ：國立政治大學教育學系

計畫主持人：邱美秀

計畫參與人員：專科畢-專任助理：侯玉玲

報 告 附 件 ：出席國際會議研究心得報告及發表論文

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 96年09月27日

Achievements and Self-Concepts in Comparison of Maths and Science: Exploring the Internal/External Frame of Reference Model Across 28 Countries

數學與科學成就與自我概念：探討28國的內外在參照架構模式

Running head: Achievements and Self-Concepts

Mei-Shiu Chiu

National Chengchi University, Taiwan

Dr. Mei-Shiu Chiu
Department of Education
National Chengchi University
64, Zhinan Rd. Sec. 2
Taipei 11605, Taiwan ROC
E-mail: chium@nccu.edu.tw
Tel: 886-2-29393091 ext 66135
Fax: 886-2-29396823

Notes on Contributor. Mei-Shiu Chiu received a B. A. and M. A. Degree in Education from National Taiwan Normal University. She completed her doctoral study at the Faculty of Education, Cambridge University. She has also been a teacher in the secondary sector, and is now an assistant professor in the Department of Education, National Chengchi University, Taiwan. She has carried out a range of research on the interactions between cognition, affect and creativity in mathematics learning in relation to teaching, assessment, and large-scale international educational tests.

Acknowledgements. This research was supported by the National Science Council, Taiwan (NSC 95-2522-S-004–001).

Note 1. Portions of this article were presented at the 31st Conference of the International Group for the Psychology of Mathematics Education, Seoul National University, Seoul, Korea, 8-13 July, 2007.

Note 2. **計畫成果自評**部份： 本研究報告內容與原計畫相符程度高，理應達預期目標，目前此篇文章投稿狀況如下：Chiu, M. S. (in submission). Achievements and self-concepts in comparison of maths and science: Exploring the internal/external frame of reference model across 28 countries. (*Educational Research and Evaluation*, submission on March 27, 2007; re-submission (in response to a request for major revision) on July 19, 2007)

摘要

Marsh (1986)提出內外在參照架構模式，以探討不同學科(例如數學與語文)間看似矛盾的現象：(1)高數學成就會產生正向的數學自我概念，高語文成就會產生正向語文自我概念；(2)高數學成就會產生負向的語文自我概念，高語文成就會產生負向的數學自我概念。本研究比較數學與科學二學科，此二學科可能被視為互補、而非完全相對。此「學科互補」的觀點顯然不同於此模式長久以來的「學科相對」觀點。本研究分析TIMSS 2003的28國八年級學生資料((*N* = 144,069)，結果顯示：不同國家的學生有不同程度的「互補」與「相對」觀點。此研究結果顯示有需要同時將「互補」與「相對」的觀點納入內外在參照架構模式中。

關鍵字：數學、科學、成就、自我概念、內外在參照架構模式

ABSTRACT

Marsh's (1986) Internal/External Frame of Reference (I/E) model posits a paradoxical pattern of relations between the contrasting school subjects, e.g., maths and verbal. (1) High maths achievement contributes to positive maths self-concept and high verbal achievement contributes to positive verbal self-concept. (2) High maths achievement contributes to negative verbal self-concept and high verbal achievement contributes to negative maths concept. An extension of the I/E model is made in comparison of maths and science, which may be viewed by students as contrasting or supplementary. The 'supplementary' perspective is a deviation from the 'contrasting' perspective that long underpins the I/E model. The results of exploring the I/E model for maths and science across 28 countries (*N* = 144,069) participating in the TIMSS 2003 study revealed that students from different countries adopted the contrasting and supplementary perspectives with varying degrees. The findings suggest the incorporation of both contrasting and supplementary perspectives into the I/E model.

Keywords: mathematics, science, achievement, self-concept, Internal/External Frame of Reference Model

INTRODUCTION

The educators of maths and science have long researched self-concepts for their respective subjects. Research results reveal that maths and science self-concepts have a consistent relationship with the corresponding achievements (e.g., Hannula, 2002; Lee & Brophy, 1996; Pietsch et al., 2003; Tuan et al., 2005; Whitebread & Chiu, 2004). These two lines of research however fail to make reference to each other across subjects. The aims of both science and maths education in constructivist curricula emphasise the links between different subjects of knowledge. The links will deepen students' understanding of knowledge and cultivate their abilities to apply knowledge in real life (e.g., Department for Education and Employment in the UK, 2000; Ministry of Education in Taiwan, 2006). Maths and science kills are supplementary in nature. Students are very likely to link their maths and science skills and then form their maths and science self-concepts. The detailed cross-subject effects of achievements on self-concepts are unknown for maths and science.

Marsh's (1986) Internal/External Frame of Reference (I/E) model proposes a clear study framework to predict the effects of achievements on self-concepts across subjects. The I/E model can be viewed as a general framework to explain this cross-subject effect and can serve as a reference framework for the present investigation. Marsh and Hau's (2004) study has successfully generalised the I/E model across 26 countries by analysing the data from the Programme for International Student Assessment (PISA) study of 2000. The present investigation will analyse the data from the Trends in International Maths and Science Study (TIMSS) of 2003, compiled by the International Association for the Evaluation of Educational Achievement (IEA, 2005). With reference to Marhs and Hau's research methods, the effects of achievements on self-concepts in comparison of maths and science were examined across 28 countries.

The Internal/External Frame of Reference (The I/E model)

The I/E model is intended for the explanation of high correlations between maths and verbal achievements but low or zero correlations between maths and verbal self-concepts (Marsh, 1986, 1989, 1990, 1993; Marsh & Hau, 2004). The theory posits a paradoxical pattern of relations between maths and verbal self-concepts and maths and verbal achievements: (1) Based on external/social comparison, high maths achievement contributes to positive maths self-concept and high verbal achievement contributes to positive verbal self-concept. (2) Based on internal comparison, high maths achievement contributes to negative verbal self-concept and high verbal achievement contributes to negative maths concept (cf. Diagram A in Figure 1). Marsh based his theory on the hierarchical or multidimensional structure of academic self-concepts (Marsh & Hattie, 1996; Shavelson et al., 1976; Yeung & Lee, 1999). The structure comprises a general self-concept on a higher level that incorporates within it two aspects at a lower level: a non-academic self-concept and an academic one. For instance, 'I'm popular' is a non-academic self-concept and 'I'm good at science' is an academic self-concept for a subject. The process of internal (ipsative-like) comparison brings students the awareness of their own relative strengths across subjects. The reference frame is within the person. The second basis of the I/E model is social comparison theory. Individuals exercise a process of external (normative-like) comparison with some standards or others in the settings. For instance, students in a high-achieving school will have a low self-concept and vice versa (Manger & Eikeland, 1997; Marsh, 1987; Marsh et al., 2000). The reference frame is external norms.

INSERT FIGURE 1 ABOUT HERE

Extensions of the I/E Model

The I/E model has been examined by intensive studies and most results supported the predictions of the model. At the early stage, the I/E model focuses on the two distinct school subjects: maths and verbal (Bong, 1998; Marsh et al., 1988; Williams & Montogmery, 1995). As a later development, Marsh et al. (2001) extended the I/E model to three subjects: Chinese, English and maths. Marsh and Yeung (2001) extended the model to the subjects of maths, Spanish and verbal. It is suggested that the I/E model will be best supported for distinct school subjects, e.g., maths vs. verbal, and native vs. nonnative languages. Marsh et al. (2001) analysed longitudinal data and found that the I/E model were quite stable over time. The I/E model is supported for not only general students but also gifted (Lee et al., 2000; Plucker & Stocking, 2001) and able students (Williams & Montgomery, 1995).

The I/E model was widely examined for students from different countries: Canada (Marsh et al., 1988), Germany (Moeller & Koller, 2001), Hong Kong (Lee et al., 2000; Marsh et al., 2001), Norway (Skaalvik & Rankin, 1995), the United Arab Emirates (Abu-Hilal & Bahri, 2000) and the United States (Bong, 1998; Plucker & Stocking, 2001; Williams & Montogmery, 1995). Marsh and Hau (2004) verified the I/E model by investigating the PISA 2000 data across 26 countries. The I/E model was also validated when either intrinsic motivations or success expectations were incorporated into the original I/E model for maths and verbal (Skaalvik & Rankin, 1995).

Most studies on the I/E model used the statistical method of structural equation modeling (SEM). SEM can test the assumptions and hypotheses posited by the I/E model all at the same time and appears to be an appropriate method. Multiple regressions are another statistical method to examine the I/E model but take several steps to explore all the concerns of the I/E model (e.g., Williams & Montogmery, 1995; Abu-Hilal & Bahri, 2000). Moeller and Koller (2001) conducted three experiments to induce university students' internal and external comparisons; the students' response patterns were consistent with the predictions of the I/E model.

Methodological Issues for Cross-Cultural Studies

Any research on cross-cultural comparison inevitably faces the problems of equivalence and bias (Van de Vijver & Leung, 2000). Theoretically, equivalent scores have a low degree of bias in cross-cultural comparison. Van de Vijver and Leung (1997) further distinguish these two terms: Bias is concerned about the issue of validity, which should be dealt with in the procedure of instrument development (e.g., item translation) and data collection (e.g., sampling procedures). Equivalence focuses on measurement issues and includes three levels of equivalence: construct (or structural) equivalence, measurement unit equivalence, and scalar equivalence (or full score comparability). For a secondary analysis based on an existing database here, appropriate statistical methods should be performed to improve score equivalence. Multilevel analysis can model the cluster structure of the sampling procedure and take account of the variances from different levels (Bryk & Raudenbush, 2002; Raudenbush et al., 2004; Snijders & Bosker, 1999). SEM can increase construct equivalence (Van de Vijver & Leung, 1997) and multigroup SEM is suitable for examining the external validity of a theory-based model (Sue, 1999). Scalar equivalence can be improved by the inclusion of the intercept term for each measured variable in the measurement model of SEM (Hair et al., 2006). The inclusion of intercept terms will allow the estimation of the means among data from different countries in multigroup SEM. Thus the meaning of each construct for each country can be taken into account.

Caution should still be exercised as ‘Western bias’ (Van de Vijver & Leung, 2000, p. 38) may inevitably remain however stringent and advanced the statistical methods are to improve equivalence. The I/E model worked as the study framework here and was used to examine the data from different countries of diverse cultures. The I/E model is basically developed from the Western culture, though positively evidenced across a number of countries. The construct of self-concepts may mean different things for students from different countries. The TIMSS survey was developed based on the psychological measurement paradigms of the Western.

The Present Study

The I/E model is well-supported by research in comparison of maths and verbal-related subjects, e.g., Chinese, English, Norwegian, Spanish, and history. The I/E model is likely to be best supported for the school subjects that are perceived by students as distinct or contrasting constructs, e.g., maths and verbal. However, maths and science can be viewed as contrasting or supplementary. The contents of maths and science show a structure whereby maths is one of the bases of science. Maths is the study of pattern and logic (Burton, 1994) and can be a powerful tool to model a wide range of domains of world knowledge, especially science. The concept of maths as providing the building blocks of science can also be found in the test content designed by the TIMSS 2003 study: Science includes the subject areas of earth science, life science, physics, chemistry and environmental science, which consist of a wide range of domains of knowledge. Maths includes the areas of algebra, data, fractions/numbers, geometry, and measurement, which comprise the knowledge of signs, images, logic and formula, all of which are widely utilised throughout each area of science. Maths therefore may be perceived as the basis of science and science as a later development or as a combination of multiple abilities. In addition, students from different countries may have different constructs of maths, science and their relations.

The present study will focus on the subjects of maths and science. The original I/E model provides a study framework that posits a stringent hypothesis for distinct school subjects, e.g., maths and verbal (Marsh & Hau, 2004): There is (1) a smaller and negative (-) path leading from maths (verbal) achievements to verbal (maths) self-concepts, controlling for (2) a substantial and positive (++) path leading from maths (verbal) achievements to maths (verbal) self-concepts (Diagram A in Figure 1). In addition, also as controlling, there is a correlation between maths and verbal achievements because of a general factor (i.e., intelligence) in different school subjects; there is also a likely correlation between maths and verbal self-concepts. Based on past studies on affective issues in maths and science education, it is sensible to hypothesise a positive path from maths achievement to maths self-concept and a positive path from science achievement to science self-concept (Diagram B in Figure 1). There are however no predictions for the cross-subject paths because maths and science are likely to be viewed as supplementary or contrasting. There is also a correlation between maths and science achievements and between maths and science self-concepts. In addition, the measurement model includes the intercept terms for each measured variable in order to take into account scalar equivalence.

The statistical methods and data presentation formats here were similar to those in Marsh and Hau's (2004) study for comparison. Both studies examined the I/E model using the statistical procedure of SEM based on data from the international databases of educational assessment. Except for the different study frameworks for the I/E model (Figure 1), the major differences between Marsh and Hau's study and the present study include: (1) PISA 2000 vs. TIMSS 2003 databases; (2) without construct intercept terms vs. with construct intercept terms for each measured variable, (3) listwise deletion vs. multiple imputation approaches to missing data, (4) single level vs. multilevel SEM. The procedures for (3) and (4) will be stated in the section of *Statistical Analysis*.

METHOD

Data Resource and Sample

The data analysed here were taken from the database of the TIMSS 2003 study. The database includes maths and science achievements of students in the eighth grade from 47 countries. All the 47 countries had the data of student maths self-concept. There are however only 28 countries with the data of student science self-concept and so the total sample was 144,069 students nested in the 5,021 classes of the 28 countries (Table 3).

Indicators

Four kinds of indicators were used to explore the I/E model. Maths and science achievements are the predictors, while maths and science self-concepts are the outcome variables. The achievement scores are computed based on the item response theory, which estimates students' latent ability of the subjects, with a score range from 5.00 to 973.01 here.

(1) *Maths achievement*, including achievements in algebra (*Mean* (*M*) = 447.97, *Standard Deviation* (*SD*) = 125.69), data (*M* = 453.22, *SD* = 120.62), fractions/numbers (*M* = 448.63, *SD* = 125.58), geometry (*M* = 446.62, *SD* = 127.30), and measurement (*M* = 447.81, *SD* = 121.90) (TIMSS-variables bsmalg01, bsmdap01, bsmfns01, bsmgeo01, and bsmmea01, respectively).

(2) *Science achievement*, including achievements in earth science (*M* = 455.77, *SD* = 122.30), life science (*M* = 460.11, *SD* = 122.44), physics (*M* = 454.42, *SD* = 126.24), chemistry (*M* = 455.59, *SD* = 121.62) and environmental science (*M* = 464.85, *SD* = 120.40) (TIMSS-variables bsseas01, bsslis01, bssphy01, bssche01, and bsseri01 respectively).

The items of self-concept used a four-point rating scale ranging from 1 (*agree a lot*) to 4 (*disagree a lot*). The scores were reverse-coded here, so that the larger numbers represented the more positive self-concepts and vice versa.

(3) *Maths self-concept*, referring to students' confidence about learning maths. The items are:

I usually do well in maths (*M* = 2.99, *SD* = .84) (TIMSS-variable bsbmtwel).

I learn things quickly in maths (*M* = 2.79, *SD* = .92) (TIMSS-variable bsbmtqky).

(4) *Science self-concept*, referring to students' confidence about learning science, with the same item content and scaling method as those for maths self-concept, except where 'science' is the school subject. The items are:

I usually do well in science (*M* = 3.13, *SD* = .84) (TIMSS-variable bsbstwel).

I learn things quickly in science (*M* = 2.96, *SD* = .89) (TIMSS-variable bsbstqky).

The above four items of self-concepts were chosen because the item contents were relevant to those used in the past studies on academic self-concepts. Exploratory factor analyses (EFA) and confirmatory factor analyses (CFA) for all the students revealed a clear structure of two factors: maths and science self-concepts. The fit indices of CFA indicated a good fit and no factor loadings were below .50. (The test results are not presented here, but similar results can be found in Table 1.) There were two negatively worded items related to science and maths self-concepts respectively, which failed to form a desirable structure with the two factors. The negatively worded items therefore were not included for analysis.

Statistical Analysis

The major statistical methods used here were SEM with the software of LISREL 8.72 (Du Toit & Du Toit, 2001; Joreskog & Sorbom, 2001, 2005). Before the major analyses, we have to solve three problems: in an order of (1) missing data, (2) uncommon measurement scale, and (3) the cluster structure of the sampling design in the TIMSS study.

(1) Missing data: Missing data are an inevitable problem for studies relying on databases (Trautwein, 2007). The proportions of missing data were 3.6% and 4.0 % for the two items of maths self-concept and 2.9% and 3.3% for those of science self-concept. The data were missing at random. There were no missing data for the other indicators. The procedure of multiple imputation (Olinsky, Chen, & Harlow, 2003; Schafer, 1997) was implemented using the LISREL software. A new data set was generated by the procedure. The new data set remained the same as the original one except for the estimated values that replaced the missing data.

(2) Uncommon measurement scale: The standardisation procedure can facilitate the SEM procedure and data presentation formats without at the expense of reliable and valid estimation of parameters. (Analyses of raw score and standardised scores resulted in the same parameter estimates here.) This procedure is especially necessary for the present data where the indicators have large differences in the score ranges between achievements and self-concepts. For instance, if raw scores were used for the multigroup SEM procedure, the common metric standardised solutions for all the achievement scores would be larger than 100.00 and those for self-concepts would be smaller than 1.00. Van de Vijver and Leung (1997) warn that standardisation for each cultural group will eliminate the differences between different cultures. Here, all the values of the 14 indicators respectively were transformed into standardised z scores ($M = 0$, $SD = 1$) based on the data of all the students from the 28 countries; the relative position of each student in the total group remained the same for each indicator. Therefore, there will be no elimination of cultural differences in analysis.

(3) The cluster structure of the sampling design: Multilevel analysis can take account of the variances from different levels of data. TIMSS data were collected by the sampling structure of three levels: students, classes and countries. For all students from all the countries as a whole, three-level (students, classes and countries) SEM was performed to examine the invariance of the a priori model across the three levels. For each country as a whole, two-level (students and classes) SEM was performed. Detailed procedures for multilevel SEM can be found in Du Toit and Du Toit (2001) and Stapleton (2002).

The indices of model fit used here include χ^2, comparative fit index (CFI), non-normed fit Index (NNFI) and root mean square error of approximation (RMSEA). The values of χ^2 are normally very large because of the large sample sizes here and so cannot be viewed as a suitable index of overall model fit (Bollen & Long, 1993; Browne & Cudeck, 1993). The values of CFI and NNFI larger than .90 and that of RMSEA smaller than .08 indicate a reasonable fit; RMSEA is also an appropriate fit index for analysis of large sample sizes (Hair et al., 1998, 2006; Schumacker & Lomax, 1996).

Multigroup SEM is a recommended procedure for cultural comparison. There are different degrees of invariance between groups (Hair et al., 2006; Marsh & Hau, 2004). (1) Loose cross-validation: The same model is imposed on different groups. (2) Partial cross-validation: Some parameters are freely estimated for each group and some parameters are constrained to be equal across groups. (3) Tight cross-validation: All parameter estimates are invariant across groups. Loose cross-validation is the basis as each group is examined their individual goodness-of-fit to a model; no comparison fit is actually made between groups. Tight cross-validation is a difficult aim for cultural-comparison analysis as countries are normally different in nature. As for partial cross-validation, error variances are viewed as the least important and firstly allowed to be freely estimated across groups. Path parameters are normally the study focus and viewed as the most important. The next important one is factor loadings. Some other parameters are of middle degrees of importance, e.g., construct intercept terms, factor covariance, factor variances, and covariance between latent variables. Similar procedures had been successfully used in Marsh and Hau's study and were used in the present study.

RESULTS AND DISCUSSION

The present results were compared with those in Marsh and Hau's (2004) study. Therefore, the discussions about the comparison and related issues are reported here.

Total Group Solutions

All the data from the 144,069 students as a total group were examined their fit to the proposed framework for the I/E model (Diagram B in Figure 1). As can be seen in Table 2, the solutions for both single-level analysis (TG1) and multilevel analysis (TG2) revealed that the data were well fit to the proposed framework: CFI and NNFI were all .99; RMSEA was .05 for Model TG1 and .04 for Model TG2. Table 1 shows that the major parameter estimates in a standardised format from the single-level SEM were the same as those from the multilevel SEM. All the parameter estimates showed that the a priori model was proper, without the problems of multicollinearity (Marsh et al., 2004) and offending estimates (Hair et al., 1998). All the factor loadings were above .50 and below 1.00. The path coefficients were consistent with the predictions of the original I/E model (Diagram A in Figure 1): (1) negative paths from maths achievement to science self-concept (-.48) and from science achievement to maths self-concept (-.15); (2) positive paths from maths achievement to maths self-concept (.13) and from science achievement to science self-concept (.29). The values of uniqueness were all significant at the .05 level. All the standard errors of each parameter estimates were .00 or .01. (The standard errors are not presented, as the parameter estimates presented in Table 1 are standardised solutions.) The correlations between the latent constructs (i.e. the standardised solutions shown in the variance-covariance matrix) were all below .90. But correlations between maths and science achievement is slightly high (.85). In comparison with the Table 1 in Marsh and

Hau's study, the correlation between maths and verbal achievement was .78 for the single-level total group solution. It is not surprising that there is a much higher relationship between maths and science achievements. In Marsh and Hau's study, the correlation between maths and verbal self-concepts is .11 and here the correlation is .50. The high correlation between maths and science self-concepts violated the starting point of the original I/E model: high correlations between maths and verbal achievements but zero to weak correlations between maths and verbal self-concepts. An odd phenomenon is that here there were negative correlations between maths achievement and science self-concept (-.23) and between science achievement and science self-concept (-.11). In Marsh and Hau's study, these two correlations were zero. This phenomenon will be discussed further in the section of *Single-Group Solutions for Selected Countries*.

INSERT TABLE 1 AND TABLE 2 ABOUT HERE

Multigroup Solutions

Multigroup SEM can examine the degrees of goodness-of-fit for the proposed model across the 28 countries. Multigroup CFA is to examine whether the measurement model is invariant across the 28 countries. The fit indices generally indicate a poor fit (Table 2). Only the less restricted models, MG1-MG3 and MG6-MG8, show a slightly good fit in terms of the values of CFI and NNFI (above .90). According to Hair et al. (2006), CFI and NNFI larger than .92 indicates a good fit but RMSEA (.14) is still too large to make the claim of a good fit given the complicated model and the large sample size here. The poor goodness-of-fit implies that there is no need to report the parameter estimates generated from the multigroup SEM procedure. But some parameter estimates were still presented in Table 1 and Table 3 here for comparison with those reported in the Table 1 and Table 3 of Marsh and Hau's study.

INSERT TABLE 3 ABOUT HERE

In Table 1 here, the parameter estimates obtained by the multigroup analysis for Model MG11 (cf. Table 2) were the same as those obtained by the other two total-group analyses for Models TG1 and TG2. In Table 3 here, the total group solution is based on the analysis for Model MG11. The path coefficients were consistent with the predictions of the I/E model: The paths from maths achievement to maths self-concept (.13) and from science achievement to science self-concept (.45) were positive; in addition, the cross-subject paths were negative: -.48 from maths achievement to science self-concept and -.15 from science achievement to maths self-concept. Based on the path coefficients in the Table 3s of both studies, the proportions of the I/E-model-fit countries were .77 (= 20/26) in Marsh and Hau's study and .46 (= 13/28) here.

An explanation for the low proportion of the I/E-model-fit countries here can be referred to the origins of the I/E model. Comparing the two Table 3s in both studies, in Marsh and Hau's study the mean factor correlation between maths and verbal for all the countries is .76 (*SD* = .05), and here is .71 (*SD* = .09) between maths and science. The difference is small. There is however a large difference in the factor correlations of the two self-concepts between the two studies. In Marsh and Hau's study the factor correlations between two self-concepts were between -.20 and .52 (*M* = .06, *SD* = .17) across countries, and here between .11 and .59 (*M* = .41, *SD* = .13). It is also quite clear in Marsh and Hau's study that the I/E model fit occurred for the 20 countries with smaller factor correlations (-.20 to .18) between maths and verbal self-concept; for the six countries of poor I/E-model fit, the correlations were positive and larger (.13 to .52). The high correlations between maths and science self-concepts here imply that the internal comparison between maths and science achievements/ability is likely to be relatively supplementary, instead of completely contrasting. But an interesting phenomenon is that the I/E model fit did occur even though the correlations between maths and science self-concepts were high in the 13 countries as shown in Table 3. This issue will be raised further in the section of *Single-Group Solutions for Each Country* because the solutions there were more reliable.

The results of the present multigroup analyses showed a poor I/E-model fit across the 28 countries (see fit indices in Table 2). Therefore, the outcomes obtained by the multigroup analyses here should be explained with caution. And it will be sensible to let each country say their respective stories about I/E model fit. Marsh and Hau's study had a generally good I/E-model fit across the countries and there was no need to go for each country.

Single-Group Solutions for Each Country

The data of each country were separately examined their fit to the a priori model (Diagram B in Figure 1) using the procedure of multilevel SEM at the levels of classes and students. The single-group SEM analyses for each country can be viewed as loose cross-validation. The fit indices revealed a good fit for each country (all CFI and NNFI > .92 and RMSEA < .08) (Table 4). Like those in Table 3, the factor correlations between maths and science self-concepts were generally large (.16 to .62, *M* = .42, *SD* = .12) and those between maths and science achievement were consistently large (.54 to .83, *M* = .70, *SD* = .08). In addition, all the correlations were significant. There were only nine countries showing I/E model fit. But the meaning of the I/E model can be further explored if we consider the large correlations between maths and science self-concepts for these nine countries. Marsh developed his theory of the I/E model based on the repeated findings that there were large correlations between maths and verbal achievements (.42 to .94) and small correlations between maths and verbal self-concepts (-.10 to +.19) (cf. Marsh, 1986). The present findings revealed that the I/E model still fitted to nine countries even though maths and science self-concepts were strongly correlated (.20 to .55) in the nine countries. A speculation is that students in the nine countries were representative of the students who function psychologically using the I/E model. The students in the nine countries have a clear, differentiated system of academic self-concepts in comparison of maths and science.

They clearly distinguish their relative strengths and the slight differences between their maths and science abilities, although their maths and science achievements were highly correlated and their maths and science self-concepts were also substantially correlated. In other words, students in these nine countries had a contrasting system of academic self-concepts in comparison of maths and science. On the other hand, the other 19 countries had a relatively less differentiated system of self-concepts for maths and science; they view maths and science as relatively supplementary, instead of completely contrasting. According to Table 4, there were in total nine patterns of cross-subject path coefficients. To facilitate interpretations, the nine patterns may be roughly organised from completely contrasting to completely supplementary (Figure 2). The Pattern 1 (I/E model fit) countries formed the largest group, which explains why the solutions of the total-group analyses (Table 1) were consistent with the predictions of the I/E model. The diversity of the patterns can be part of the reasons for the poor model fit obtained by the multigroup analyses (cf. the fit indices in Table 2).

INSERT TABLE 4 AND FIGURE 2 ABOUT HERE

An interesting question is: Are there characteristics common to the countries within a specific pattern? Given the small sample sizes in each pattern, any answers will be quite misleading. But the nine countries of I/E model fit did reflect a specific kind of culture. Their official language is English or they are historically influenced by English culture. Of course this is a tentative answer because Australia did not belong to Pattern 1, though it did belong to the adjacent Pattern 2. ‘Self’ or ‘self-concept’ appears to be an important heritage from the Western (Singelis, 2000); therefore a highly differentiated system of ‘academic self-concepts’ will be developed by the students within the culture. They can distinguish the slight difference between their maths and science abilities. The sensitivity to the relative strengths through the process of internal comparison enables students to focus on their specific strength in the early years and along the life span. Perhaps it is even tentative to explain the supplementary perspective given little literature and the few cases of countries here. Based on comparison with the contrasting perspective, one speculation is that students with supplementary perspectives are likely to pay efforts to integrate their diverse abilities and pay attention to both their relative strengths and weaknesses. In other words, the supplementary perspective implies a less differentiated system of academic self-concepts. Markus and Kitayama (1991) proposed a dichotomy between the independent and interdependent views on self. The original I/E model bases the theory on an independent or contrasting view across subjects. Perhaps some cultures tend to cultivate an interdependent or supplementary view across subjects. This tendency brings about a relatively interdependent or supplementary system of academic self-concepts. Future research can address this issue further.

Single-Group Solutions for Selected Countries

Some methodological limitations of the present study will be discussed with help from detailed information of single-group solutions for some selected countries. In Table 3, there were some path coefficients larger than 1.00, which were offending estimates (Hair et al., 1998) and implies a poor fit of Model MG9 to four countries: Taiwan, Norway, Singapore and the United States. The parameter estimates based on the single-group SEM at the class and student levels for the four countries revealed desirable estimates (Table 5). No parameter estimates were larger than 1.00 and all factor loadings were larger than .50.

INSERT TABLE 5 ABOUT HERE

The reliability coefficients of internal consistency (Chronbach's α) for each factor is presented for each country in Table 3. Tunisia and Egypt had some α coefficients below .50. The factor loadings based on multilevel SEM procedure (Table 5) also revealed a similar trend: In Tunisia, the measurement components of achievements were not acceptable, with all factor loadings below .50. The results of an EFA of the ten achievement indicators for Tunisia however revealed a structure of two factors: maths and science achievements. (The results are not presented here.) In Egypt, the measurement components of self-concepts were not acceptable, with most factor loadings below .50. The results of an EFA of the four self-concept indicators revealed only one factor. In other words, for students in Egypt, the four items of maths and science self-concepts had similar meanings. The low reliability coefficients and low factor loadings suggest that the parameter estimates based on the data of Egypt and Tunisia should be explained with caution.

An issue raised here is how many items should be included in order to form a reliable construct. Many homogeneous items placed together in a questionnaire will form a reliable construct. The TIMSS 2003 study obviously took a different approach. The self-concept items were placed with a number of items of different constructs, though maths and science items were placed in different sections of the questionnaire. There were only two self-concept items worded for maths and science respectively in a positive manner. The other two negatively worded items were unreliable across countries. Similar phenomenon was found in the PISA 2003 study, as reported in Marsh and Hau's study for the items of verbal self-concept; one negatively worded item of verbal was deleted eventually in their study. Marsh (1994) raised similar issue in analysing the data from the National Educational Longitudinal Survey of 1988. He suggested that positively and negatively worded items have different constructs in measurement. He successfully dealt with the problem by the approach of setting correlated errors between the items. Here, this approach was tried for not only negatively worded items but also for the same wording items of maths and science self-concepts; but all the tries failed. Now a trend in psychological measurement is to include as few items as possible for one construct and still maintain psychometric quality. The major reason is that long questionnaires take more time to complete, have higher refusal rates, and have more missing data than short questionnaires

(Russell et al., 2004; Stanton, 2000; Stanton et al., 2002). International databases, like TIMSS and PISA, have endeavored to develop high-quality items for a number of important constructs in education and psychology; they also adjusted the items to different countries. Analyses of the items taken from these databases however revealed that it seems to be an impossible dream to have ‘a valid and reliable common psychological ruler’ for any specific construct across countries. If measurement invariance cannot be achieved, any claims based on results of secondary analysis across countries will be questionable. A reverse thought is: Can we accept the use of one single valid item for one psychological construct? This issue is beyond the scope of the present study but can be addressed by studies on educational and psychological measurement.

The final discussion will focus on the patterns of the correlations between maths achievement and maths self-concept, between maths achievement and science self-concept, between science achievement and maths self-concept, and between science achievement and science self-concept. In the Table 1 (the variance-covariance matrix) of Marsh and Hau’s study, these correlations were all zero for total-group and multigroup solutions. In Table 1 here, the correlations were negative (-.04 to -.23), except for those between maths achievement and maths self-concept. On the other hand, the single-group solutions for selected countries shown in Table 5 here revealed positive relationships, most of which were substantial (.03 to .61). Marsh and Hau’s study did not report the solutions based on single-group analyses. A speculation is that there is a critical country factor ‘moderating’ (Baron & Kenny, 1986) these relationships. A likely country factor is the big-fish-little-pond effect: Students in high-achieving groups (e.g., schools) tend to have low self-concept and vice versa (Manger & Eikeland, 1997; Marsh et al, 2000; Marsh & Parker, 1984; Marsh & Rowe, 1996). As Leung’s (2002) study indicated, the high-achieving Asian countries participating in the TIMSS study confessed a lower self-concept in learning maths than some other countries. Although multilevel SEM was performed for the total group here, country factors in fact failed to be successfully identified. The procedure for multilevel analysis can be used to identify country moderators (Bryk & Raudenbush, 2002; Muthen & Muthen, 1998-2006; Snijders & Bosker. 1999; Raudenbush et al., 2004). There are also likely some mediators intervening in the process of internal and external comparisons; some effective mediators may include perceived school status (Marsh et al., 2000), interest, goals, values and anxiety (Ho et al, 2000; Pintrich & De Groot, 1990; Turner et al., 1998; Wigfield & Eccles, 2000). Combining mediators and moderators into the I/E model will complicate the model but may be a likely way to elaborate the theory.

CONCLUSION

The I/E model has advanced to an era of a general framework for all different school subjects. The exploration of the I/E model in comparison of maths and science broadened the scope of the model. Maths and science are two highly related school subjects in nature, and may be viewed by students as contrasting or supplementary in terms of internal comparison. The ‘supplementary’ perspective is a deviation from the ‘contrasting’ perspective that traditionally underpins the I/E model. The results of exploring the I/E model for maths and science across 28 countries revealed that students from different countries adopted the contrasting and supplementary perspectives with varying degrees. The findings suggest the incorporation of both contrasting and supplementary perspectives into the I/E model or into the multidimensional structure of academic self-concepts. The incorporation of cultural, contextual and pedagogical factors into the I/E model is likely to further increase the benefits of the model for educational practices.

The TIMSS 2003 study had endeavoured to reduce construct, method and item bias in the process of instrument development and data collection. The present secondary analysis applied the procedure of multiple imputation to deal with missing data; score equivalence was improved by multilevel SEM, multigroup SEM, and the inclusion of construct intercept terms for each measured variable. There are however some limitations that dictate the need to interpret the results with caution. The problems of measurement and sampling errors were not fully solved. The proposed study framework failed to address the issue of the reciprocal relations between self-concepts and achievements; the reciprocal or causal relations can be investigated with a number of waves of data. The organisations of the large-scale international educational tests may need to further improve measurement reliability, validity and equivalence for cross-cultural comparison. The use of negatively worded items should be further examined not only based on the theories of psychological measurement but also based on the theories of cultural psychology.

REFERENCES

Abu-Hilal, M. M., & Bahri, T. M. (2000). Self concept: The generalisability of research on the SDQ, Marsh/Shavelson model and I/E frame of reference model to the United Arab Emirates students. *Social Behaviour and Personality, 28,* 309-322.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Newbury Park, CA: Sage.

Bong, M. (1998). Tests of the internal/external frames of reference model with subject-specific academic self-efficacy and frame-specific academic self-concepts. *Journal of Educational Psychology, 90,* 102-110.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Bryk, A. S., & Raudenbush, S. W. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

Department for Education and Employment (2000). *National curriculum*. Retrieved 27 September 2001 from the World Wide Web: www.nc.uk.net

Du Toit, M., & Du Toit, S. H. C. (2001). *Interactive LISREL: User's guide*. Lincolnwood, IL: Scientific Software International, Inc.

Hair, J. F. Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

Hair, J. F. Jr., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Prentice-Hall.

Hannula, M. S. (2002). Attitude towards mathematics: Emotions, expectations and values. *Educational Studies in Mathematics, 49*, 25-46.

Ho, H., Senturk, D., Lam, A. G., Zimmer, J. M., Hong, S., Okamoto, Y., Chiu, S., Nakazawa, Y., & Wang, C. (2000) The affective and cognitive dimensions of math anxiety: A cross-national study. *Journal for Research in Mathematics Education, 31*, 362-379.

International Association for the Evaluation of Educational Achievement. (2005). *TIMSS 2003 user guide for the international database*. Chestnut Hill, MA: TIMSS & PIRLS International Study Centre.

Joreskog, K. G., & Sorbom, D. (2001). *LISREL 8: User's reference guide.* Lincolnwood: IL: Scientific Software International.

Joreskog, K. G., & Sorbom, D. (2005). *LISREL 8.72* [*computer software*]. Lincolnwood: IL: Scientific Software International, Inc.

Lee, M. F., Yeung, A. S., Low, R., & Jin, P. (2000). Academic self-concept of talented students: Factor structure and applicability of the internal/external frame of reference model. *Journal for the Education of the Gifted, 23*, 343-367.

Lee, O., & Brophy, J. (1996). Motivational patterns observed in sixth-grade science classrooms. *Journal of Research in Science Teaching, 33*, 303-318.

Leung, F. K. S. (2002). Behind the high achievement of East Asian students. *Educational Research and Evaluation, 8*, 87-108.

Manger, T., & Eikeland, O. J. (1997). The effect of social comparison on mathematics self-concept. *Scandinavian Journal of Psychology, 38*, 237-241.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224-253.

Marsh, H. W. (1986). Verbal and maths self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23*, 129-149.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology, 79*, 280-295.

Marsh, H. W. (1989). Sex differences in the development of verbal and mathematics constructs: The high school and beyond study. *American Educational Research Journal, 26*, 191-225.

Marsh, H. W. (1990). A multidimensional, hierarchical self-concept: Theoretical and empirical justification. *Educational Psychology Review, 2*, 77-172.

Marsh, H. W. (1993). Academic self-concept: Theory measurement and research. In J. Suls (Ed.), *Psychological perspectives on the self* (Vol. 4, pp. 59-98). Hillsdale, NJ: Erlbaum.

Marsh, H. W. (1994). Using the national longitudinal study of 1988 to evaluate theoretical models of self-concept: The Self-Description Questionnaire. *Journal of Educational Psychology, 86,* 439-456.

Marsh, H. W., & Hattie, J. (1996). Theoretical perspectives on the structure of self-concept. In B. A. Bracken (Ed.), *Handbook of self-concept* (pp. 38-90). New York: Wiley.

Marsh, H. W., & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalisability of the Internal/External frame of reference predictions across 26 countries. *Journal of Educational Psychology, 96*, 56-67.

Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology, 47*, 213-231.

Marsh, H. W., & Rowe, K. J. (1996). The negative effects of school-average ability on academic self-concept: An application of multilevel modelling. *Australian Journal of Education, 40*, 65-87.

Marsh, H. W., & Yeung, A. S. (2001). An extension of the internal/external frame of reference model: A response to Bong (1998). *Multivariate Behavioural Research, 36,* 389-420.

Marsh, H. W., Byrne, B. M., & Shavelson, R. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology, 80,* 366-380.

Marsh, H. W., Dowson, M., Pietsch, J., & Walker, R. (2004). Why multicollinearity matters: A re-examination of relations between self-efficacy, self-concept and achievement. *Journal of Educational Psychology, 96,* 518-522.

Marsh, H. W., Kong, C. K., & Hau, K. T. (2000). Longitudinal multilevel models of the big-fish-little-pond effect on academic self-concept: Counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology, 78*, 337-349.

Marsh, H. W., Kong, C. K., & Hau, K. T. (2001). Extension of the internal/external frame of reference model of self-concept formation: Importance of native and nonnative languages for Chinese students. *Journal of Educational Psychology, 93,* 543-553.

Ministry of Education in Taiwan (2006). *The national curriculum*. Retrieved 14 February 2007 from the World Wide Web: http://teach.eje.edu.tw/9CC/index.php. (in Chinese)

Moeller, J., & Koller, O. (2001). Dimensional comparisons: An experimental approach to the internal/external frame of reference model. *Journal of Educational Psychology, 93,* 826-835.

Muthen, L. K., & Muthen, B. O. (1998-2006). *Mplus user's guide*. Los Angeles, CA: Muthen & Muthen.

Olinsky, A., Chen, S., & Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modelling. *European Journal of Operational Research*, *151*, 53-79.

Pietsch, J., Walker, R., & Chapman, E. (2003). The relationship among self-concept, self-efficacy and performance in mathematics during secondary school. *Journal of Educational Psychology, 95,* 589-603.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82,* 33-40.

Plucker, J. A., & Stocking, V. B. (2001). Looking outside and inside: Self-concept development of gifted adolescents. *Exceptional Children, 67,* 535-548.

Raudenbush, S., Bryk, A., Cheong, Y. F., Congdon, R., & du Toit, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Russell, S. S., Spitzmuller, C., Lin, L. F., Stantion, J. M., Smith, P. C., & Ironson, G. H. (2004). Shorter can also be better: the abridged job in general scale. *Educational and Psychological Measurement, 64,* 878-893.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Erlbaum.

Shavelson, R., Hubner, J., & Stanton, G. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46,* 407-441.

Singelis, T. M. (2000). Some thoughts on the future of cross-cultural social psychology. *Journal of Cross-Cultural Psychology, 31*, 76-91.

Skaalvik, E. M., & Rankin, R. J. (1995). A test of the internal/external frame of reference model at different levels of math verbal self-perception. *American Educational Research Journal, 32,* 161-184.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modelling*. London: Sage.

Stanton, J. M. (2000). Empirical distributions of correlations as a tool for scale reduction. *Behavior Research Methods, Instruments, and Computers, 32,* 403-406.

Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology, 55*, 167-194.

Stapleton, L. M. (2002). The incorporation of sample weights into multilevel structural equation models. *Structural Equation Modelling, 9*, 475-502.

Sue, S. (1999). Science, ethnicity and bias: Where have we gone wrong? *American Psychologist, 54,* 1070-1077.

Trautwein, U. (2007). The homework-achievement relation reconsidered: Differentiating homework time, homework frequency, and homework effort. *Learning and Instruction, 17*, 372-388.

Tuan, H. L., Chin, C. C., & Shieh, S. H. (2005). The development of a questionnaire to measure students' motivation towards science learning. *International Journal of Science Education, 27*, 639-654.

Turner, J. C., Thorpe, P. K., & Meyer, D. K. (1998). Students' reports of motivation and negative affect: A theoretical and empirical analysis. *Journal of Educational Psychology, 90*, 758-771.

Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research.* Thousand Oaks, CA: Sage

Van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology, 31*, 33-51.

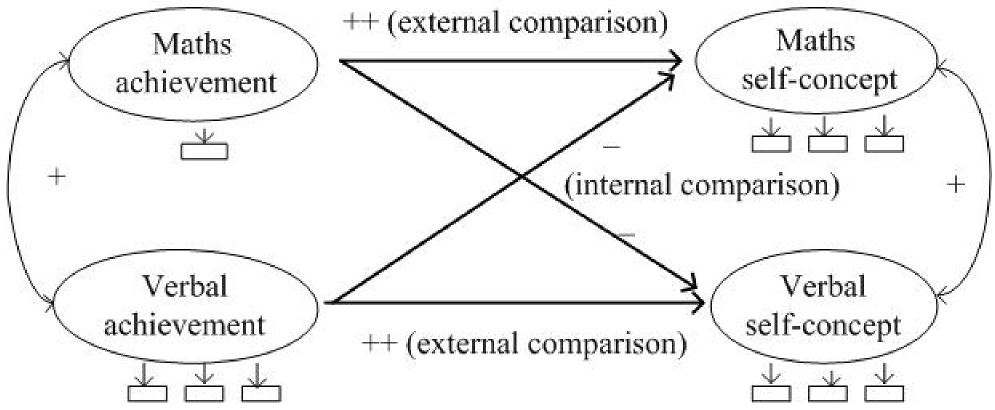
Whitebread, D. & Chiu, M. (2004). Patterns of children's emotional responses to mathematical problem-solving. *Research in Mathematics Education, 6*, 129-153.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*, 68-81.

Williams, J. E., & Montogmery, D. (1995). Using frame of reference theory to understand the self-concept of academically able students. *Journal for the Education of the Gifted, 18,* 400-409.

Yeung, A. S., & Lee, F. L. (1999). Self-concept of high school students in China: Confirmatory factor analysis of longitudinal data. *Educational and Psychological Measurement, 59,* 431-450.

A. The study framework for the I/E model in Marsh and Hau's (2004) study



B. The present study framework for the I/E model

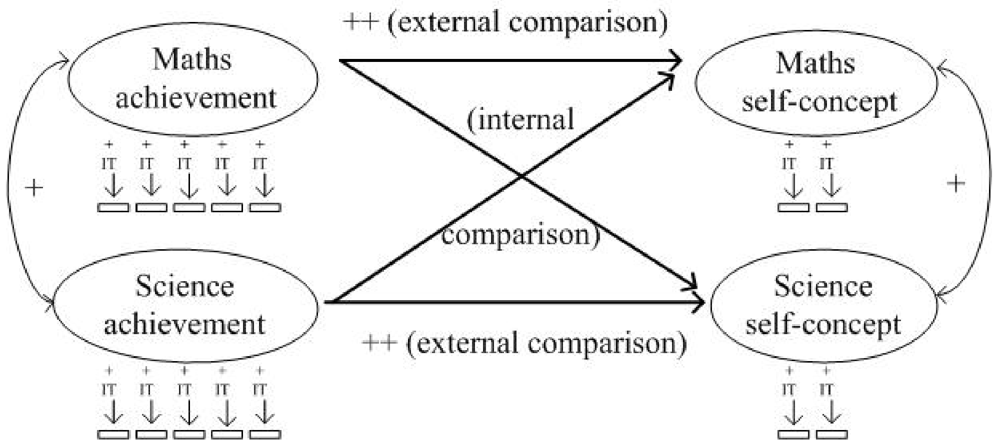


Figure 1. The I/E model predicts (1) a smaller and negative (-) path leading from maths (verbal) achievements to verbal (maths) self-concepts, controlling for (2) a substantial and positive (++) path leading from maths (verbal) achievements to maths (verbal) self-concepts. The present study provides no predictions for the cross-subject paths because maths and science are likely to be viewed as contrasting or supplementary. The construct intercept term (IT) is included to be estimated for each measured variable.

Patterns	1 Contrasting	2	3	4	5	6	7	8	9 Supplementary
Path from MA to SS	-	-	ns.	-	+	ns.	ns.	+	+
Path from SA to MS	-	ns.	-	+	-	ns.	+	ns.	+
Countries	Botswana Hong Kong Malaysia New Zealand Philippines Singapore South Africa United States England	Australia Chile Syrian Arab Tunisia Egypt	Bahrain	Taiwan Israel	Morocco Scotland	Italy Korea Saudi Arabia	Jordan Norway	Ghana Iran Japan	Palestinian Nat'l Auth
Numbers of countries	9	5	1	2	2	3	2	3	1

Figure 2. According to the path coefficients shown in Table 4, there are nine patterns of cross-subject paths, which are roughly organised from completely contrasting to completely supplementary. The countries in Pattern 1 are I/E model fit. '+' = positive path coefficient; '-' = negative path coefficient; ns. = non-significant path coefficient.

Table 1

Parameter Estimates for Total-Group Solution (Model TG1), 3-Level Total-Group Solution (Model TG2) and Multigroup Solution (MG11)

	Student-level total-group solution					3-level total-group solution				Multigroup solution			
Factor	MA	SA	MS	SS	Uniq.	MA	SA	MS	SS	MA	SA	MS	SS
Factor loading													
MA1	.95				.10	.95				.95			
MA2	.91				.17	.91				.91			
MA3	.97				.06	.97				.97			
MA4	.93				.14	.93				.93			
MA5	.94				.11	.94				.94			
SA1		.94			.11		.94				.94		
SA2		.96			.08		.96				.96		
SA3		.95			.10		.95				.95		
SA4		.94			.62		.94				.94		
SA5		.94			.61		.94				.94		
MS1			.79		.37			.79				.79	
MS2			.69		.52			.69				.69	
SS1				.86	.26				.86				.86
SS2				.68	.54				.68				.68
Path coefficient													
MA													
SA													
MS	.13	-.15				.13	-.15			.13	-.15		
SS	-.48	.29				-.48	.29			-.48	.29		
Variance-covariance													
MA	1.00					1.00				1.00			
SA	.85	1.00				.85	1.00			.85	1.00		
MS	.01	-.04	1.00			.01	-.04	1.00		.01	-.04	1.00	
SS	-.23	-.11	.50	1.00		-.23	-.11	.50	1.00	-.23	-.11	.50	1.00

Note. All parameter estimates are common metric standardised solutions. The total-group solution is based on Model TG1 (Table2), the 3-level total-group solution is based on Model TG2, and the multigroup solution is based on MG11 (with all the parameters invariant across the 28 countries). MA = maths achievement; SA = science achievement; MS = maths self-concept; SS = Science self-concept; Uniq. = uniqueness.

Table 2

Goodness-of-Fit Indices for the I/E Model Fit to the Total Group and Multiple (Country) Groups

Model	χ^2	*df*	CFI	NNFI	RMSEA	Model description
Total group SEM						
TG1	23078.81	71	.99	.99	.05	SEM: Single (student) level
TG2	69236.42	309	.99	.99	.04	SEM: 3-level
Multigroup CFA						
MG1	175108.80	2258	.94	.93	.14	CFA Inv. = none; Free =FL, FV, IT, Uniq.
MG2	199451.63	2528	.93	.93	.14	CFA Inv. = FL; Free = FV, IT, Uniq.
MG3	231129.83	2798	.91	.92	.14	CFA Inv. = FL, FV; Free = IT, Uniq.
MG4	572104.98	3176	.79	.83	.21	CFA Inv. = FL, FV, IT; Free = Uniq.
MG5	581407.37	3284	.78	.83	.21	CFA Inv. = FL, FV, IT, Uniq. (total invariance)
Multigroup SEM						
MG6	202477.68	2591	.93	.93	.14	SEM Inv. = PC; Free = FL, FV, IT, Uniq.
MG7	229624.98	2771	.92	.92	.14	SEM Inv. = PC, FL; Free = FV, IT, Uniq.
MG8	231129.83	2798	.91	.92	.14	SEM Inv. = PC, FL, FV; Free = IT, Uniq.
MG9	552808.19	3068	.80	.83	.21	SEM Inv. = FL, FV, IT; Free = PC, Uniq.
MG10	572104.98	3176	.79	.83	.21	SEM Inv. = PC, FL, FV, IT; Free = Uniq.
MG11	581407.37	3284	.78	.83	.21	SEM Inv. = PC, FL, FV, IT, Uniq. (total invariance)

Note. *N* = 144,069. Model TG1 (Table 1 shows parameter estimates) examines whether the I/E model was fit to the total group. Model TG2 (Table 1 shows parameter estimates) also uses the total-group sample and is a three-level (countries, classes and students) SEM. Models MG1-MG11 examine whether the I/E model was fit to each of the 28 countries by setting some combination of invariant and free parameters across the countries. TG = total group; MG = multiple group (or multigroup); CFA = confirmatory factor analysis; SEM = structural equation model; CFI = comparative fit index; NNFI = non-normed fit Index; RMSEA = root mean square error of approximation; Inv. = invariant; Free = free estimation; FL = factor loadings; FV = factor variance-covariances; IT = construct intercept terms; PC = path coefficients; Uniq. = uniqueness;.

Table 3

Reliability Estimates, Goodness-of-Fit Indices and Selected Parameter Estimates Based on Multigroup Solutions for Each Country

Country	Coun-try Code	*N*	Reliability MA *α*	SA *α*	MS *α*	SS *α*	Factor co. MA-SA	MS-SS	Path coefficient MA to MS	MA to SS	SA to MS	SA to SS	I/E
Total		144,069	.97	.88	.71	.74	.85	.50	.13	-.48	-.15	.29	
1. Australia	36	4,791	.96	.75	.77	.78	.77	.41	.86	-.07	-.30	.14	Yes
2. Bahrain	48	4,199	.94	.65	.55	.57	.66	.45	.07	-.41	.21	.45	
3. Botswana	72	5,150	.93	.73	.53	.55	.64	.34	.37	-.26	-.22	.26	Yes
4. Chile	152	6,377	.96	.74	.73	.72	.72	.19	.46	-.36	-.06	.39	Yes
5. Taiwan	158	5,379	.98	.86	.81	.80	.83	.49	.63	-.26	-1.13	-.25	
6. Palestinian Nat'l Auth	275	5,357	.95	.78	.56	.54	.72	.58	.02	-.39	.31	.61	
7. Ghana	288	5,100	.85	.86	.53	.59	.64	.51	-.08	-.33	-.08	-.07	
8. Hong Kong	344	4,972	.96	.81	.75	.73	.76	.39	-.05	-.94	-.36	.57	
9. Iran, Islamic Rep. of	364	4,942	.95	.66	.71	.67	.64	.45	.31	-.16	.24	.48	
10. Israel	376	4,318	.96	.75	.67	.67	.72	.32	.59	-.23	.00	.68	
11. Italy	380	4,278	.96	.73	.83	.76	.71	.40	.75	-.09	-.25	.26	Yes
12. Japan	392	4,856	.97	.80	.70	.72	.72	.59	-.15	-.83	-.60	.03	
13. Jordan	400	4,489	.95	.78	.56	.52	.75	.61	-.11	-.43	.54	.68	
14. Korea	410	5,309	.97	.78	.79	.76	.78	.53	.16	-.78	-.59	.13	Yes
15. Malaysia	458	5,314	.96	.70	.64	.64	.72	.50	.37	-.58	-.42	.16	Yes
16. Morocco	504	2,943	.93	.61	.66	.63	.49	.45	.28	-.16	-.22	.05	
17. New Zealand	554	3,801	.96	.73	.76	.77	.76	.34	.93	-.11	-.43	.11	Yes
18. Norway	578	4,133	.95	.67	.77	.75	.69	.41	1.12	-.04	-.17	.32	
19. Philippines	608	6,917	.95	.80	.48	.54	.71	.49	.22	-.21	-.21	.12	Yes
20. Saudi Arabia	682	4,295	.89	.61	.56	.59	.57	.48	-.10	-.43	.11	.32	
21. Singapore	702	6,018	.97	.89	.79	.79	.85	.21	.26	-1.06	-.45	1.05	Yes
22. South Africa	710	8,952	.93	.91	.60	.57	.76	.37	.14	-.26	-.20	.08	Yes
23. Syrian Arab	760	4,895	.92	.59	.58	.55	.55	.36	-.13	-.66	.14	.36	
24. Tunisia	788	4,931	.90	.43	.64	.60	.52	.11	.62	-.44	-.40	-.12	
25. Egypt	818	7,095	.97	.84	.45	.49	.77	.53	.28	-.07	-.09	.15	Yes
26. United States	840	8,912	.97	.81	.79	.79	.80	.19	1.04	-.23	-.46	.50	Yes
27. England	926	2,830	.96	.80	.72	.79	.80	.31	.69	-.08	-.29	.18	Yes
28. Scotland	927	3,516	.96	.74	.71	.81	.78	.44	.72	-.05	-.13	.50	
M			.95	.74	.67	.67	.71	.41	.37	-.35	-.20	.29	
SD			.03	.10	.11	.10	.09	.13	.38	.28	.33	.28	

Note. The parameter estimates of path coefficients are common metric standardised solutions across groups (countries). The 28 countries were placed into the procedure of multigroup SEM in the order of the Country Code reported in the TIMSS database. All the estimates of factor correlations and path coefficients are significant at the .05 level, except for those between -.06 and .06 are not significant at the .05 level and others are significant. The solution of the total sample is based on Model MG11. The factor correlations of each country are obtained by analysis on Model MG2 (Table 2). Path coefficients are based on Model MG9. Factor co. = factor correlation; I/E = the I/E model; MA = maths achievement; SA = science achievement; MS = maths self-concept; SS = Science self-concept; Yes = fit to the predictions of the I/E model in terms of path coefficients.

Table 4

Goodness-of-Fit Indices and Selected Parameter Estimates Based on Multilevel Solutions for Each Country

	Factor co.		Path coefficient					Goodness of fit			
Country	MA -SA	MS- SS	MA to MS	MA to SS	SA to MS	SA to SS	I/E	χ^2(190)	CFI	NNFI	RM -SE A
1. Australia	.75	.41	.54	-.07	-.01 ns	.37		2329.52	.99	.99	.05
2. Bahrain	.67	.46	.50	-.01 ns	-.05	.23		1974.64	.99	.99	.05
3. Botswana	.65	.37	.27	-.16	-.23	.31	Yes	1518.16	.99	.99	.04
4. Chile	.71	.20	.36	-.18	-.02 ns	.28		3326.22	.99	.99	.06
5. Taiwan	.80	.47	.56	-.08	.04	.44		2835.54	.99	.99	.06
6. Palestinian Nat'l Auth	.72	.59	.42	.06	.09	.37		1340.86	.99	.99	.03
7. Ghana	.69	.52	.19	.04	.04 ns	.34		1624.08	.99	.99	.04
8. Hong Kong, SAR	.72	.38	.47	-.28	-.08	.46	Yes	2780.58	.99	.99	.05
9. Iran, Islamic Rep. Of	.64	.46	.41	.09	.02 ns	.22		1205.33	.99	.99	.03
10. Israel	.70	.33	.41	-.11	.03	.44		1845.88	.99	.99	.05
11. Italy	.69	.39	.50	00 ns	.01 ns	.29		2069.19	.99	.99	.05
12. Japan	.70	.55	.56	.05	.02 ns	.42		2169.51	.99	.99	.05
13. Jordan	.74	.62	.34	.03 ns	.07	.27		1278.41	.99	.99	.04
14. Korea, Rep.of	.74	.51	.65	.00 ns	.00 ns	.51		1639.31	.99	1.00	.04
15. Malaysia	.69	.48	.41	-.22	-.08	.56	Yes	2936.37	.99	.99	.05
16. Morocco	.50	.49	.25	.05	-.06	.16		1357.34	.98	.98	.05
17. New Zealand	.74	.34	.62	-.10	-.18	.41	Yes	2109.36	.99	.99	.05
18. Norway	.69	.41	.58	-.02 ns	.05	.34		1899.21	.99	.99	.05
19. Philippines	.71	.51	.31	-.15	-.24	.25	Yes	1844.90	.99	.99	.04
20. Saudi Arabia	.60	.49	.31	.01 ns	.01 ns	.22		1085.75	.99	.99	.04
21. Singapore	.83	.20	.79	-.47	-.44	.66	Yes	4089.16	.99	.99	.06
22. South Africa	.78	.38	.26	-.15	-.17	.23	Yes	2185.98	.99	.99	.03
23. Syrian Arab	.54	.37	.27	-.09	-.02 ns	.16		1500.67	.99	.99	.04
24. Tunisia	.52	.12	.40	-.15	-.01 ns	.34		2689.59	.97	.97	.05
25. Egypt	.77	.55	.36	-.10	-.04 ns	.32		2134.06	.99	.99	.04
26. United States	.77	.55	.59	-.09	-.18	.41	Yes	4526.84	.99	.99	.06
27. England	.78	.32	.53	-.16	-.20	.47	Yes	1180.31	.99	.99	.04
28. Scotland	.76	.44	.53	.04	-.11	.44		1159.34	.99	.99	.04
M	.70	.42	.44	-.08	-.06	.35					
SD	.08	.12	.14	.12	.12	.12					

Note. The parameter estimates are common metric standardised solutions. The data of each country were separately examined their goodness-of-fit to the a priori model (Diagram B in Figure 1) using multilevel structural equation modelling at the class and student levels. ALL estimates of factor correlations and path coefficients are significant at the .05 level except for those indicated by 'ns' (non-significant). Factor co. = Factor correlation. I/E = the I/E model. CFI = comparative fit index; NNFI = non-normed fit Index; RMSEA = root mean square error of approximation. MA = maths achievement; SA = science achievement; MS = maths self-concept; SS = Science self-concept; Yes = fit to the predictions of the I/E model in terms of path coefficients..

Table 5

Parameter Estimates of Multilevel Solutions for Taiwan, Norway, Singapore, Tunisia, Egypt and the United States

	Taiwan				Norway				Singapore				Tunisia				Egypt				United States			
Factor	MA	SA	MS	SS	MA	SA	MS	SS	MA	SA	MS	SS	MA	SA	MS	SS	MA	SA	MS	SS	MA	SA	MS	SS
Factor loading																								
MA1	.82				.55				.62				.44				.80				.58			
MA2	.60				.57				.55				.44				.67				.58			
MA3	.80				.54				.60				.44				.71				.64			
MA4	.83				.49				.58				.41				.81				.53			
MA5	.75				.52				.61				.45				.74				.62			
SA1		.56				.54				.68				.39				.87				.61		
SA2		.58				.55				.66				.42				.76				.64		
SA3		.55				.49				.55				.49				.80				.55		
SA4		.85				.60				.82				.28				.83				.76		
SA5		.59				.61				.75				.36				.78				.72		
MS1			.93				.81				.86				.62				.42				.84	
MS2			.76				.75				.77				.79				.48				.77	
SS1				.89				.78				.81				.47				.41				.83
SS2				.70				.63				.74				.70				.53				.80
Path coefficient																								
MA																								
SA																								
MS	.56	.04			.58	.05			.79	-.44			.40	-.01			.36	-.04			.59	-.18		
SS	-.08	.44			-.02	.34			-.47	.66			-.15	.34			-.10	.32			-.09	.41		
Variance-covariance																								
MA	1.00				1.00				1.00				1.00				1.00				1.00			
SA	.80	1.00			.69	1.00			.83	1.00			.52	1.00			.77	1.00			.77	1.00		
MS	.59	.49	1.00		.61	.44	1.00		.42	.21	1.00		.39	.19	1.00		.33	.24	1.00		.45	.27	1.00	
SS	.27	.38	.47	1.00	.21	.33	.41	1.00	.08	.28	.20	1.00	.03	.27	.12	1.00	.15	.24	.55	1.00	.22	.34	.18	1.00

Note. All parameter estimates are common metric standardised solutions. The data of each country were separately examined their goodness-of-fit to the a priori model (see Figure 1) using the procedure of multilevel structural equation modelling at the levels of classes and students (the same as the procedure used in Table 4). MA = maths achievement; SA = science achievement; MS = maths self-concept; SS = Science self-concept.

行政院國家科學委員會補助國內專家學者出席國際學術會議報告

96 年 9 月 7 日

報告人姓名	邱美秀	服務機構及職稱	國立政治大學教育學系 助理教授
時間 會議地點	2007 年 8 月 28 至 9 月 1 日 布達佩斯 (匈牙利)	本會核定補助文號	NSC 95-2522-S-004–001
會議名稱	(中文)第十二屆歐洲學習與教學研究會議 (英文) 12th EUROPEAN CONFERENCE FOR RESEARCH ON LEARNING AND INSTRUCTION		
發表論文題目	(中文)數學深度學習之後設認知知識量表：以多重方法學發展 (英文) The scale of levels of meta-cognitive knowledge in achieving deep approaches to mathematics learning: Developed by a mixed methodology.		

報告內容

一、參加會議經過

28 日註冊、開幕、論文發表會

29 日發表論文、參與論文發表會與演講

30 日參與論文發表會與演講

31 日參與論文發表會與演講

9 月 1 日早上參與論文發表會，下午至機場搭機返回台灣

二、與會心得

1、此學術研討會歷史會歐洲極大型教育學術研討會，與會學者來自世界各地、歐洲學者為最大宗。包含知名資深、新生代、與研究生超過千人。台灣有約十位學者(含教授與研究生)參與並發表論文。

2、活動內容豐富、論文水準高、熱心的學者們提供建設性的意見以促進此領域之學術研究。

3、此團體主要由一群國際委員們規畫，學術活動規畫周詳、豐富，社群活動的規畫亦很貼心。

4、整體而言，這是一個高水準的國際學術研究團體與研討會。

三、考察參觀活動(無是項活動者省略) 無

四、建議

此會議較少台灣學者與研究生參與，而此研討會可視為歐洲極重要的教育組織，事實上其參與者來自全球。如果台灣想擴大於歐洲的參與度，此為重要的學術會議。

五、攜回資料名稱及內容

ABSTRACTS FOR 12th BIENNIAL CONFERENCE FOR RESEARCH ON LEARNING AND INSTRUCTION. BUDAPEST, HUNGARY (CD 片一張)