An Empirical Study of Gender Difference in the Relationship between Self-Concept and Mathematics Achievement in a Cross-Cultural Context

Jianjun Wang*
California State University, USA

The reciprocal relationship between mathematics achievement and self-concept is examined in this study to expand the existing knowledge to a cross-cultural setting. Based on analyses of educational data two years before and after Hong Kong’s sovereignty switch in 1997, this investigation shows a weak reciprocal relationship among the eighth-grade students across gender categories. With the introduction of a new policy to promote Chinese instruction, changes in the perceived importance of English have been assessed by education stakeholders, and the results are used to indicate a latent factor of “English push” behind the learning process. In the context of cross-cultural transition, gender differences have been found in path coefficients toward mathematics achievement and self-concept. Limited by the four-year research period, one may speculate whether it would take longer to demonstrate a large effect size in the reciprocal relationship.

Mathematics is an important school subject taught in many countries. Students’ mathematics achievement and positive self-concept are quality indicators of workforce preparation in the global marketplace (Evans, 2005). In an attempt to model the relationship between student achievement and self-concept, Shavelson, Hubner, and Stanton (1976) proposed a multifaceted, hierarchical structure that split general self-concept into academic and nonacademic components. Similar to the “chicken–egg” puzzle, academic self-concept may cause changes in academic achievement, or vice versa. To reflect this linkage, a theoretical model postulates

*Department of Advanced Educational Studies, California State University, Bakersfield, 9001 Stockdale Highway, Bakersfield, CA 93311, USA. Email: jwang@csub.edu
that “academic self-concept and academic achievement are reciprocally related and mutually reinforcing: improved academic self-concept will lead to greater achievement, and greater achievement will lead to improved academic self-concept” (Marsh, Hau, & Kong, 2002, p. 729).

The reciprocal relation model needs to be reconfirmed in cross-cultural settings to enhance its generalisability. According to Marsh et al. (2002), “In academic self-concept research, support for the main theoretical models has been based largely on responses by students from Western countries, particularly English-speaking students in Australia, Canada, and the United States” (p. 728). In addition, Byrnes, Hong, and Xing (1997) examined Chinese students’ performance on an American mathematics test, and found a significant interaction between gender and culture factors: Despite a large gender difference in the test performance of the American students, no such gap was found in their Chinese counterpart. Marsh et al. (2002) concurred: “Previous research suggests that Chinese students differ from Western students in ways that may be relevant to how they construct their self-concepts” (p. 728).

Before and after being returned to China on July 1, 1997, Hong Kong participated in the Third International Mathematics and Science Study (TIMSS) in 1995 and a repeat of TIMSS (TIMSS-R) in 1999, respectively. Two months after the sovereignty switch, the new Hong Kong government put forth a policy that required instruction in Chinese in most secondary schools (Dan, 1997). As a result, there may have been changes to the perceived importance of English among education stakeholders, and empirical data have been gathered in the TIMSS and TIMSS-R projects to indicate a latent factor of “English push” behind the learning process. The purpose of this investigation is to examine gender difference in the relationship between mathematics achievement and self-concept under the context of policy change during the political transition. To date, except for a study by Wang (2004), no one has conducted this type of investigations using the TIMSS and TIMSS-R databases. Wang’s (2004) study did not consider gender difference in the reciprocal relation modelling. To fill this void in the research literature, empirical findings from this investigation may not only enrich existing knowledge of the reciprocal relation model, but also articulate the outcomes of language policy that are particularly relevant to the period of cross-cultural transition in Hong Kong.

**Literature Review**

Since its inception in 1959, the International Association for the Evaluation of Educational Achievement (IEA) has conducted a series of comparative studies to provide information about educational achievement and learning context. The TIMSS project launched in 1995 was the largest of these studies, and involved 41 nations with dual foci on mathematics and science education. TIMSS-R in 1999 repeated the TIMSS investigation at the eighth grade in 38 nations.

Like their predecessors, the TIMSS and TIMSS-R projects have demonstrated two major applications: besides comparing student performance across nations,
these projects also support independent investigations within each country. Permission has been granted to each country to add special items to the international instrument during data collection (Martin, Gregory, & Stemler, 2000). As a result, some researchers have set aside its comparative nature, and chose to focus their TIMSS and TIMSS-R data analyses within a single nation (e.g., Casey, Nuttall, & Pezaris, 2001; Hosenfeld, Koller, & Baumert, 1999; Wester & Henriksson, 2000).

**Uniqueness of the Hong Kong Study**

Hong Kong has a unique cross-cultural setting that accommodates both English and Chinese instruction in secondary education. Sweeting (1997) noted: “Language issues have infected educational practice and policy from the very earliest days of Hong Kong’s existence” (p. 181). Instruction in Chinese was a new policy strictly enforced in the transition period between the TIMSS and TIMSS-R projects. Lau (1998, p. 1) recollected:

> Although the decision to use mother tongue as the medium of instruction was taken in the final years of British colonial rule, the fact that it is being pushed vigorously by the SAR government [the new Hong Kong government] inevitably gives the impression that this is because Hong Kong is now under Chinese rule.

In contrast to Hong Kong’s peaceful reunification with China, the language change has generated strong discontent in Hong Kong. According to Lau (1998, p. 1):

> The emotions stirred up by the Hong Kong Special Administrative Region (SAR) government’s decision to force local secondary schools to use Chinese as the medium of instruction has taken many people by surprise. School principals, teachers, students and parents are passive and seldom express their views with the vehemence seen in the past few months.

Holliday and Holliday (2003) noted that “language is an important cultural factor when comparatively assessing students who speak, read, write, and listen using entirely different communication system” (p. 252). In Hong Kong, the native Chinese language is widely used in daily conversations while English is primarily confined to professional communications (Evans, 2000). Reducing the role of English in secondary education has profound implications for Hong Kong’s future as an international business center (Bray, 1997). Because secondary education is the final phase of compulsory education, disentangling the factor of English push is a well-justified approach to analysing the ultimate impact of the new language policy at the student level during the cross-cultural transition.

Although some teachers might take a while to abandon mixed-code instruction, the policy announcement could have immediately influenced the perceived importance of the English language by mothers, peers, and student themselves (Wang, 2004). Students had a chance to report the perspectives of their mothers, peers, and themselves during the TIMSS and TIMSS-R data collection, and these variables have been coded on a scale of “strongly agree” to “strongly disagree” to indicate the level of English push placed upon students. The data codebooks for TIMSS and TIMSS-R are available for download from http://www.timss.org.
Meanwhile, “To investigate how students think of their abilities in mathematics, TIMSS created an index of students’ self-concept in mathematics (SCM)” (Mullis et al., 2001, ch. 4, p. 129). In addition, “The Third International Mathematics and Science Study (TIMSS) offers a unique opportunity to examine some of the issues related to gender differences in mathematics achievement” (Fierros, 1999, p. 1). Besides the general reports of gender difference in mathematics achievement for all participating nations (Beaton et al., 1996; Mullis, Martin, Gonzalez et al., 2000), more in-depth investigations have been conducted on gender difference within specific countries (Casey et al., 2001; Cheng & Seng, 2001; Ercikam, McCreith, & Lapointe, 2005; Hosenfeld et al., 1999; Wester & Henriksson, 2000).

In short, the period of TIMSS and TIMSS-R data-gathering coincided with the era of historical transition in Hong Kong. No other large-scale data have been gathered in this period to support an empirical comparison two years before and after the sovereignty handover. Besides including indicators of the English push factor particularly sensitive to the language policy change, these two projects are also pertinent to investigations of mathematics achievement (Beaton et al., 1996; Mullis, Martin, Fierros et al., 2000), self-concept (Mullis et al., 2001; Wang, 2004), and gender difference (Casey et al., 2001; Cheng & Seng, 2001; Ercikam et al., 2005; Fierros, 1999; Hosenfeld et al., 1999; Wester & Henriksson, 2000). Built on these important features of TIMSS and TIMSS-R, this study is designed to examine gender difference in the reciprocal relationship between student self-concept and mathematics achievement in the context of policy change during the sovereignty handover.

Factors Connected to the Reciprocal Relation Model

Byrne (1984) stressed the motivational properties of academic self-concept: changes in self-concept can lead to changes in academic achievement. Conversely, academic self-concept may emerge as a consequence of academic achievement (Lau, Yeung, & Jin, 1998). Marsh et al. (2002) reported that “the results of previous research provide general support for a reciprocal effects model” (p. 729).

More specifically, however, comparative studies have disagreed on details of the reciprocal relationship (see Wilkins, 2003). For instance, Wilkins, Zembylas, and Travers (2002) analysed the TIMSS international data, and reported a positive relationship between self-concept and academic achievement for 16 different countries. Meanwhile, Kifer’s (2002) analysis suggested that many of the highest-performing countries had some of the weakest student beliefs in his/her self-ability.

Marsh et al. (2002) recognised the importance of Hong Kong for verifying the reciprocal relationship, and attempted to differentiate instructional practice between Chinese and English schools in this cross-cultural context. Yip, Tsang, and Cheung (2003, p. 303) observed:

A problem with the design of the Marsh et al. study, which might affect the validity of data interpretation, is that many of the so-called English-medium schools used Chinese
or mixed code for instruction, so only a small number of the EMI [English Medium Instruction] were truly English medium.

Because elementary schools in Hong Kong generally adopted Chinese as their instructional language (Evans, 2002), teachers might have chosen to incorporate Chinese lectures in English-medium schools to reduce difficulty arising from the language switch in secondary education. In addition, mixed-code instruction was related to the fact that not all teachers were capable of teaching proficiently in English. Consequently, “a substantial proportion of schools which claimed to teach in English actually taught either in Cantonese or in mixed code” (Bray, 1997, p. 162). Fok (2001, p. 6) further clarified this dilemma from parental perspectives:

In Hong Kong, a mastery of English means greater ability for earning money. This is why many parents want their children to have good English. But the problem is that the use of English as the medium of instruction has deprived children of the chance of a good self-concept education because not many children are able to cope with English.

The language issue may filter through to mathematics education in different ways. One speculation is that students could learn more effectively when taught in their mother tongue (see Cummins, 1996; Garcia, 1993), and the improvement of achievement might lead to enhancement of positive attitude, as suggested by the reciprocal relation model (Marsh, Byrne, & Yeung, 1999). Others may argue that “English language high schools tended to be the most prestigious and the most academically selective high schools in Hong Kong” (Marsh et al., 2002, p. 732). Thus, students might have better self-concepts in these settings, which could counterbalance the “cultural ambivalence” arising from late immersion into the second language instruction (e.g., Krashen, 1997; Ogbu, 1992, 1999). Given this cross-cultural complexity, Tao (1994) recollected: “The medium of instruction has been a very controversial issue in Hong Kong for decades” (p. 323).

Meanwhile, gender difference is an important factor in self-concept studies across different cultures. Singelis (2000) concurred that a study of self (male or female) has been the primary reason that many disciplines are increasingly embracing cross-cultural perspectives. Wong, Lam, and Ho (2002) further noted: “The traditional Chinese culture is male-oriented culture” (p. 830). Marsh’s (1993) analysis in this area concluded that gender differences in specific domains of self-concept were typically consistent with gender stereotypes. Thus, Chinese cultural heritage seemed to support the stereotypic differentiation between boys and girls. However, Hong Kong was ruled by Britain for the most part of the last century. “This western outlook has been translated to equitable distribution of educational and employment opportunities for both men and women” (Wong et al., 2002, p. 830), which might narrow the gender gap.

In summary, Wilgenbusch and Merrell (1999) reported: “Gender differences and gender issues in self-concept have been a topic of considerable interest for at least two decades. However, the empirical base for assertions in this area is still somewhat limited” (p. 103). To enrich existing knowledge, extensive discussions have been presented in this section to justify the value of investigating gender difference in the
reciprocal relationship between self-concept and mathematics achievement in a cross-cultural context. Features of the TIMSS and TIMSS-R projects have been reviewed along with the research literature on reciprocal relation modelling to support this secondary data analysis at this historical junction of Hong Kong’s sovereignty handover.

Research Questions

Mathematics is a core school subject. “Overall, schools see themselves as competing with other schools for pupils and the key determinant of their success is their academic achievements in high-status subjects” (Morris & Chan, 1997, p. 252). Given the focus on student learning within a school setting, variance in education outcomes should be partitioned at the student and school levels to facilitate result generalisation across school settings. Accordingly, the research questions that guide this investigation are:

1. Are there any gender differences in the proportion of mathematics score variance that has been distributed at the student and school levels?
2. Are there any gender differences in the relationship between student self-concept and mathematics achievement, given the English push factor during the political transition?
3. What are the differences and/or similarities of the empirical model before and after the political transition, as illustrated from the TIMSS and TIMSS-R data analyses?

Method

Measures of Mathematics Achievement

Scores of mathematics achievement have been gathered from the eighth-grade level in the TIMSS and TIMSS-R projects. To avoid a low response rate, each student was tested on a subset of mathematics items, and a total of five plausible scores have been imputed to represent overall student achievement (Wang, 2001). According to Gonzalez and Smith (1997), “one set of the imputed plausible scores can be considered as good as another” (p. 3). The use of multiple plausible scores has been recommended by other national projects in the U.S., such as the National Assessment of Educational Progress (Allen, Carlson, & Zelenak, 1999).

To enhance comparability between the TIMSS and TIMSS-R results, TIMSS scores have been rescaled at the eighth-grade level using the TIMSS-R procedure, and the data are available for comparison online (http://www.timss.org). In this study, variance from each of the five plausible scores is partitioned at the student and school levels, and the final result is based on an average of the findings from the five rounds of computing in each of the gender categories (Question 1). If the score variation is fairly small at the school level, then no additional school variables are needed to explain the outcome of student performance (Raudenbush & Bryk, 2002).
Other variables besides mathematics scores could have been considered in the multilevel analysis. However, those variables are either a categorical factor (i.e., gender) or ordinal indicators of the self-concept and English push factors. No statistical method is available to compute variances from these nominal or ranking data for multilevel variance partitioning. On the other hand, correlation coefficients can be computed from ranking data. On basis of the correlation matrix, those ordinal variables can be analysed using the structural equation modelling (SEM) technique (Garson, 2005). Gender can be employed as a classification variable for a multigroup SEM analysis (Muthen & Muthen, 2005). The SEM features are discussed below to reflect current research findings on statistical modelling.

Statistical Modelling

Development of individual self-concept can be dated back to Socrates’ call to “know thyself” more than 2,000 years ago (see review by Hamachek, 2000). In line with the traditional perception of self as “I” or “me”, the individual self-concept can be split into two aspects, “the self as a doer” and “the self as an object” (Hamachek, 2000; James, 1890). From the doer’s perspective, TIMSS and TIMSS-R gathered “students’ self-perceptions about usually doing well in mathematics” (Beaton et al., 1996, p. 118). Regarding the self as an object, students had a chance to express feelings of getting bored by mathematics. As explained previously, the English push factor can be indicated by students’ reports of the importance of English from the perspectives of their mothers, peers, and themselves. Based on the postulation that “changes in academic self-concept will lead to changes in subsequent academic achievement” (Marsh et al., 1999, p. 155), the reciprocal relation model incorporates the English push factor to articulate gender differences in the relationship between self-concept and mathematics achievement before and after the political transition (Question 2). Mullis, Martin, Fierros et al. (2000) noted that the TIMSS data are appropriate for an analysis of gender difference in academic and affective outcomes.

Figure 1 shows accommodation of multiple indicators to identify each of the latent factors for the reciprocal relation modeling. “Today, it is commonly accepted that multiple observed variables are preferred over a single variable in defining a latent variable” (Schumacker & Lomax, 1996, p. 55). To keep Figure 1 more readable, measurement errors considered in this investigation are not depicted along with factor loadings, path coefficients, and correlation coefficients. The use of multiple indicators permits an assessment of measurement errors in this study, and goodness-of-fit indices have been computed from the TIMSS and TIMSS-R databases for reconfirmation of the reciprocal relation model (Question 3). To facilitate interpretation of the research findings, these indicators have been scaled in such a way that a higher value represents a more positive response in each dimension. The LISREL software was employed to handle the statistical calculation.

Like any other secondary data analysis, this study is based on information from existing databases. At conclusion of this section, three limitations should be acknowledged accordingly:
1. Some of the criticisms of the TIMSS and TIMSS-R databases might be relevant to this investigation. Most of the criticisms were based on the question of comparability of high school students from substantially different education systems (e.g., Bracey, 1997, 1998; Rotberg, 1998; Wang, 2001). To stay away from this controversy, this investigation is confined to Hong Kong, and does not involve international comparisons at the 12th-grade level.

2. The variable selections are limited. Although the TIMSS and TIMSS-R researchers attempted to gather self-concept information (Mullis et al., 2001), the released items were somewhat restricted in comparison to other investigations in this area (e.g., Lau et al., 1998; Marsh et al., 2002). In addition, the English push factor was identified by student-reported importance of English learning.

Notes:

1. PS1–PS5 are plausible mathematics scores imputed from TIMSS or TIMSS-R projects under a three-parameter item response theory (IRT) model.

2. The ìdoerî and ìobjectî indicators are based on student responses about whether they can do well in mathematics and have felt bored about mathematics.

3. Indicators for the ìEnglish pushî factor are based on student responses regarding the importance of learning English perceived by self, friends, and parents.

Figure 1. A structural model of self-concept and mathematics achievement from TIMSS
from the mother, peer, and self perspectives (Figure 1). Other indicators, such as father’s English push, cannot be studied because no such data have been collected to disentangle the broad influence of the new language policy. The limited number of variables might have hindered this secondary data analysis from reporting results of the English push factor more adequately.

3. Given the fact that other contextual factors remained fairly stable during the peaceful transition in Hong Kong, English push appeared to be one of the most profound factors directly linked to the switch of instructional languages in secondary school (Bray, 1997). While this variable might have overshadowed other factors, it would be too simplistic to ignore other variables not gathered in TIMSS and TIMSS-R. Because of the variable preclusion, it is possible that other projects may shed more light on alternative interpretations to triangulate the results of this investigation.

Results

The TIMSS sample contains data from 3,630 male and 3,007 female students, and the TIMSS-R has 2,585 male and 2,534 female cases. The missing response rate was below 10% for both projects. Partition of the score variances is presented in Table 1 to examine whether the results can be generalised across different schools. The small variance ratio suggests relatively little variation in educational outcome among schools within Hong Kong’s urban setting (see Table 1).

Goodness of fit indices have been listed in Table 2 to assess the plausibility of fitting the reciprocal relation model (Figure 1) to the TIMSS and TIMSS-R databases in each gender category.

Given the large sample size, all factor loadings ($\lambda$s) and structural parameter estimates ($\beta$ and $\gamma$s) are significant at $\alpha = .05$, with $t$ values ranging from 4.9 to 146.6. Table 3 contains standardised parameter estimates for the reciprocal relation model in Figure 1 using the TIMSS and TIMSS-R databases. To avoid the apparent statistical significance due to sample size, the effect sizes of the structural relationships before and after Hong Kong’s political transition are presented in Table 4 to compare the model difference. This approach is in line with a recommendation from

<table>
<thead>
<tr>
<th>Project</th>
<th>Gender</th>
<th>PS1</th>
<th>PS2</th>
<th>PS3</th>
<th>PS4</th>
<th>PS5</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIMSS</td>
<td>Male</td>
<td>.050</td>
<td>.049</td>
<td>.051</td>
<td>.049</td>
<td>.047</td>
<td>.049</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>.047</td>
<td>.047</td>
<td>.049</td>
<td>.046</td>
<td>.047</td>
<td>.047</td>
</tr>
<tr>
<td>TIMSS-R</td>
<td>Male</td>
<td>.085</td>
<td>.085</td>
<td>.089</td>
<td>.087</td>
<td>.082</td>
<td>.085</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>.083</td>
<td>.090</td>
<td>.085</td>
<td>.087</td>
<td>.086</td>
<td>.086</td>
</tr>
</tbody>
</table>

*The ratio is computed by dividing the partitioned score variance at the school level by the corresponding variances at the student level.
the American Psychological Association to represent the real value differences in dissemination of statistical findings (Henson & Smith, 2000; Thompson, 1998).

**Discussion**

Students’ mathematics achievement and positive self-concept are important learning outcomes that attract the attention of various education stakeholders (Ma & Kishor, 1997). Understanding their relationship in a cross-cultural setting can facilitate development of general theories useful to educators around the world. Built on the existing knowledge base, this study focuses on examination of a reciprocal relation

### Table 2. Model fitting indices for the TIMSS and TIMSS-R databases*

<table>
<thead>
<tr>
<th></th>
<th>TIMSS samples</th>
<th></th>
<th>TIMSS-R samples</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Male Female</td>
<td>Total Male Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMR</td>
<td>.10 .09 .10</td>
<td>.10 .09 .11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFI</td>
<td>.98 .98 .98</td>
<td>.98 .98 .97</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* RMR = root mean square residual; GFI = goodness of fit index (Joreskog & Sorbom, 1993).

### Table 3. Parameter estimates for the reciprocal model using the TIMSS and TIMSS-R databases

<table>
<thead>
<tr>
<th></th>
<th>TIMSS</th>
<th></th>
<th>TIMSS-R</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male Female Effect size</td>
<td>Male Female Effect size</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Factor loadings</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{y1}$</td>
<td>.96 .96 .00</td>
<td>.93 .92 .01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{y2}$</td>
<td>.96 .95 .01</td>
<td>.93 .92 .01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{y3}$</td>
<td>.96 .96 .00</td>
<td>.93 .92 .01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{y4}$</td>
<td>.96 .95 .01</td>
<td>.93 .92 .01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{x1}$</td>
<td>.62 .57 .05</td>
<td>.51 .66 -.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{x2}$</td>
<td>.56 .63 -.07</td>
<td>.78 .64 .14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{x3}$</td>
<td>.64 .64 .00</td>
<td>.59 .64 -.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{x4}$</td>
<td>.57 .66 -.09</td>
<td>.63 .63 .00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{x5}$</td>
<td>.74 .78 -.04</td>
<td>.83 .83 .00</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reciprocal relation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>.11 .05 .06</td>
<td>.07 .08 -.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Path coefficients</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>.36 .23 .13</td>
<td>.34 .30 .04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>.28 .21 .07</td>
<td>.22 .28 -.06</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The reciprocal relation corresponds to the mutual causation linkage between student self-concept and mathematics achievement in Figure 1.
model during a special period of cross-cultural transition in Hong Kong. With the implementation of a new language policy to switch the medium of instruction from English to Chinese, education stakeholders might have changed their perceptions of the importance of English language in secondary school. The perception change can be employed to indicate a latent variable of English push behind the learning process across gender categories. Analysis of gender difference also provides an opportunity to examine the stability of the model-fitting indices over different sub-samples. Based on an overview of this project design, the discussion is divided into sub-topic areas, including a multilevel partition of the score variances, and interpretations of other statistical findings on the basis of the research literature and Hong Kong’s unique cross-cultural context. The model-fit indices are examined further at end of the discussion to guard against potential statistical artifact.

Multilevel Partition of the Mathematics Score Variances

Educators often consider schools as agents that shape gender inequalities in student performance (see Younger, Warrington, & Williams, 1999). Partition of the score variances in mathematics may help disentangle the school effect. In this study, the variance ratios between school and student levels have been computed for all five plausible scores in TIMSS and TIMSS-R. The ratio in Table 1 appears fairly stable across the plausible scores, which confirms equivalence of these scores according to the original TIMSS and TIMSS-R designs (Beaton et al., 1996; Mullis, Martin, Gonzalez et al., 2000).

The relatively small variance at the school level seems to agree with similar results from earlier studies by Willms and Raudenbush (1989) and Mortimore, Sammons, Stoll, Lewis, and Ecob (1988). To interpret this finding in a broad context, one may note that Hong Kong society is fairly homogeneous with Chinese accounting for 98% of the population (Poon, 1999). “Unlike the case in the UK, there is no official classification of social class in Hong Kong” (Lai, 2001, p. 115). Thus, families generally share similar educational values, and variation in school quality is unlikely to result from major differences in school location or diversity in socio-economic status. Meanwhile, because of Hong Kong’s free schooling, equity of education has been enhanced between male and female students in recent years (Post & Pong, 1998).

Wong et al. (2002) further observed, “Large-scale studies examining the effects of schooling on gender differences in academic achievements have seldom been conducted in Hong Kong” (p. 828). One of the major educational changes accompanying Hong Kong’s political transition is full implementation of the new language

<table>
<thead>
<tr>
<th>Gender</th>
<th>$\beta$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>-.04</td>
<td>-.02</td>
<td>-.06</td>
</tr>
<tr>
<td>Female</td>
<td>.03</td>
<td>.07</td>
<td>.07</td>
</tr>
</tbody>
</table>
policy that eliminates mixed-code instruction (Evans, 2000). The policy had the effect of sharpening the difference between English and Chinese schools. As a result, one may expect to see larger score variances at the school level in TIMSS-R because of the enforced language separation. Although no statistical testing has been developed to assess the significance level of the increase in the variance ratio from TIMSS to TIMSS-R (Raudenbush & Bryk, 2002), the magnitude of the difference seems substantial with the TIMSS-R results 40% higher than the corresponding findings from TIMSS (Table 1). On the other hand, because the language policy was not designed for or against a particular gender category, Table 1 shows that the effect size between male and female students has been consistently small for both TIMSS and TIMSS-R.

In summary, Evans (2005) pointed out: “Nearly 90% of the variance in students’ math scores on some tests can be predicted without knowing anything about their schools” (p. 584). Table 1 shows that school variations contributed less than 10% of the variation in student mathematics scores. This result appears to preclude the need to introduce more school-level factors into this investigation. This research outcome has been confirmed across gender categories (Table 1). In designing the international assessment for a multicultural context, proper attention was given to “crafting gender-fair test items in TIMSS [that] could have enabled the girls to compete on equitable grounds with the boys” (Cheng & Seng, 2001, p. 336). Thus, it is appropriate to analyse the TIMSS and TIMSS-R data across gender categories to reconfirm the multilevel statistical findings.

Interpretation of Additional Results in Terms of the Existing Literature

The released TIMSS and TIMSS-R reports have already documented gender difference in mathematics achievement in each nation, and no significant gender gap was found in the Hong Kong databases (Beaton et al., 1996; Mullis, Martin, Gonzalez et al., 2000). That result is reconfirmed in this study by similar factor loadings ($\lambda_{y1}$…$\lambda_{y5}$ of the equivalent plausible scores) across the gender classification (Table 3). Table 3 also shows smaller factor loadings for the TIMSS-R data; this reduction in factor loadings could have resulted from additional disturbance variables during the political transition, such as the sharpened language contrast under the new language policy, which reduced coherence of the latent variable identification (Table 3). Regarding gender difference in mathematics achievement, Cheng and Seng (2001, p. 332) recollected:

At the International Commission on Mathematics Instruction Conference in 1993, practically all the participating nations reported male advantage (Hanna, 1996). The only exception was the People’s Republic of China which reported an experiment carried out in Shanghai where secondary girls outperformed boys. (Tang, Zheng, & Wu, 1996)

It is interesting to note that there was a switch from males scoring higher in TIMSS (Beaton et al., 1996) to females scoring higher in TIMSS-R (Mullis,
Self-Concept and Achievement

Martin, Fierros et al., 2000). While the gap was not statistically significant, the effect size in Table 3 is in line with the reported switch of male and female score differences (Cheng & Seng, 2001).

The reciprocal relation model suggests that changes in achievement pattern will lead to changes in mathematical self-concept. Although the change in effect size is small, it is interesting to note that there was a switch of factor loadings ($\lambda_{y6}, \lambda_{y7}$ on the latent factor of self-concept between male and female students; see Table 3). For female students in the TIMSS study, the indicator of “usually doing well in mathematics” showed smaller than the indicator of “getting bored by mathematics” ($\lambda_{y6} < \lambda_{y7}$). The opposite was true in the TIMSS-R result (i.e., $\lambda_{y6} > \lambda_{y7}$). This pattern has been reversed for male students (Table 3). Therefore, similar to the reported better mathematics achievement of female students in urban China (Cheng & Seng, 2001), in Hong Kong there were better female scores than male scores (Mullis, Martin, Gonzalez et al., 2000) and more positive self-concept among females after the territory returned to China (see $\lambda_{y6} > \lambda_{y7}$ for the TIMSS-R result in Table 3). In line with the improvement in female mathematics education, the effect size also suggests an increase in the reciprocal relation ($\beta$) and path coefficients ($\gamma_1, \gamma_2$) for female students in TIMSS-R (Table 4).

**Interpretation of the Results in Hong Kong’s Context**

Historical trends should be examined to avoid exaggerating the effect arising solely from the political transition. Before the government handover, Hong Kong was run by the United Kingdom. Rogers, Galloway, Armstrong, and Leo (2000) pointed out a similar trend in the UK:

Data provided by the Department of Education in the U.K. indicate the variable nature of the differences between the genders in qualifications obtained at secondary-school level. Summaries of these data (Rogers 1986) indicate a general closing of the gender gap, with girls being more likely over time to take up and gain qualifications in subjects such as mathematics. The educational aspirations of males and females are also becoming increasingly similar. Indeed, current concerns in the U.K. are more likely to be directed at the low achievement of boys than of girls. (p. 80)

The TIMSS and TIMSS-R results seem to support this assessment by showing no significant gender difference in England at the eighth grade (Beaton et al., 1996; Mullis, Martin, Gonzalez et al., 2000). Hence, even though the TIMSS-R data were gathered only two years after Hong Kong’s transition, the change in factor loadings between male and female students (Table 3) does not represent a complete reverse of the established trend under UK governance. Instead, the changes seen in Table 3 seem to fit the ongoing adjustment of educational outcomes in the cross-cultural context (Cheng & Seng, 2001; Rogers et al., 2000).

Unique to Hong Kong’s cross-cultural transition was a switch of instructional language from English to Chinese in most secondary schools. In the TIMSS and TIMSS-R projects, the maternal push for English learning was surveyed via a question on the importance of doing well in English. Like the self and peer indicators of
the English push factor, students reported their mother’s push on an ordinal scale. Contributions from each indicator have been estimated concurrently using the structural equation modelling technique (Joreskog & Sorbom, 1993) with the magnitude represented by the factor loadings ($\lambda_{x1}$, $\lambda_{x2}$, $\lambda_{x3}$) in Table 3. Effects of the English push factor on self-concept are shown by the path coefficient ($\gamma_2$). Even in Hong Kong’s English schools, “one language [English] was used in school and a completely different one [Chinese] in all other places” (Chan, 2002, p. 274). Because the language concern is largely academic in nature (Evans, 2002), it is no surprise to observe a higher path coefficient toward the achievement factor (see $\gamma_1 > \gamma_2$ in Table 3).

Since learning in English “is seen by parents as offering the best prospect for their children’s future” (Hong Kong Education Commission, 1990, p. 93), parental protest against the new language policy after the transition was highly publicised (Evans, 2000). In time of crisis, girls are “more highly monitored [by parents] than boys” (Svensson, 2003, p. 300). Consequently, although no gender difference was found in the mother’s push indicator in TIMSS ($\lambda_{1x1}$), the TIMSS-R results showed that the mother’s push indicator was the only indicator having gender differences in factor loading after eruption of the parental protest (Table 3). The different factor loadings seemed to be justified by female receptiveness to additional parental pressure after Hong Kong’s transition. Besides the parental push, students were also exposed to peer influence in Chinese outside school. In general, boys are “more exposed to deviant peers than girls” (Svensson, 2003, p. 300). Unrestrained peer influence seems to have contributed to the factor loading difference between male and female students (Table 3).

Finally, it should be noted that the reciprocal link (β) between mathematics achievement and self-concept is much weaker than the path coefficients ($\gamma_1$, $\gamma_2$), and the corresponding effect size for gender difference is relatively small for all these indexes (see effect sizes for $\gamma_1$, $\gamma_2$ and β in Table 3). While the results fail to support a strong reciprocal relationship, the small gender difference agrees with other findings in the research literature. Ma and Kishor’s (1997) meta-analysis concurred: “There were no statistically significant gender differences on the self-concept–achievement relationship” (p. 101).

Model Fitness to the TIMSS and TIMSS-R Databases

Table 2 lists the model-fit indexes for each gender group using the TIMSS and TIMSS-R databases. The use of multiple model-fit indexes is recommended by Bollen (1989). In general, an appropriate model should have a small root mean square residual (RMR) and a high goodness of fit index (GFI) (Joreskog & Sorbom, 1993; Marcoulides & Schumacker, 1996). In terms of either criterion, the results in Table 2 clearly suggest strong support for the statistical model from the TIMSS and TIMSS-R databases.

Still, it is possible that these indexes might be “affected by sample size” (Sharma, 1996, p. 158). To examine stability of the model fitness across different sample
sizes, these indexes have been computed from the combined student samples across the gender classification. The combined database is approximately twice as large as the sub-data for each gender, and the RMR and GFI values remain stable (Table 2); this cross-examination indicates that the model fitness is unlikely to have resulted from statistical artifacts pertaining to a particular sample size (Sharma, 1996).

In summary, this investigation supports Marsh, Hau, and Kong’s (2002) assertion that “cross-cultural comparisons provide a valuable basis for testing theories and models [between self-concept and academic achievement]” (p. 727). On one hand, the multilevel analysis has resulted in a small portion of the score variance at the school level, which eliminated the need to incorporate additional school variables to explain student achievement difference at the school level. At the student level, the positive reciprocal relation ($\beta$) in Figure 1 seems to suggest mutual enforcement between mathematics achievement and self-concept, which is in line with the theory developed in the Western nations (Marsh et al., 1999). The small $\beta$ value characterises a weak linkage between science achievement and self-concept from the Hong Kong databases during the cross-cultural transition. Since this study is confined to an analysis of two large-scale databases separated by four years, one may speculate whether cultural changes take a longer time to show gender differences in educational outcomes. This speculation seems pertinent given the small values of factor loadings and effect sizes, as well as the changes to the factor loadings for self-concept items (Tables 3 and 4). Readers should be mindful of these limitations, and avoid a far-fetched interpretation beyond the available data. Assuming that the short time period has contributed to small effect sizes across the gender (male vs. female) and project (TIMSS vs. TIMSS-R) dimensions, this study seems to support a noteworthy effort to continue verifying these findings using new databases over a longer period in the future.

References


