Running head: Self-Concept & Achievement

An Empirical Study of Gender Difference

in the Relationship Between Self-Concept and Mathematics Achievement*

Jianjun Wang

Department of Advanced Educational Studies

California State University, Bakersfield

9001 Stockdale Highway

Bakersfield, CA 93311

Phone: (661) 664-3048

Fax: (661) 664-2016

E-mail: jwang@csub.edu

*Paper presented at the 2005 annual meeting of the American Educational Research Association (AERA), Montreal, Canada.

An Empirical Study of Gender Difference

in the Relationship Between Self-Concept and Mathematics Achievement

Abstract

In the western literature, mathematics achievement and its related student self-concept are important education outcomes reciprocally linked and mutually reinforcing. The reciprocal relation model is examined in this study to assess its generalization in a cross-cultural setting. Hong Kong is the site of choice because of its exposure to influences between British and Chinese cultures. A contextual factor of English push has been examined for male and female students behind the language switch from English to Chinese (Cantonese) in most secondary schools. The research findings are employed to compare features of the reciprocal model two years *before* and *after* Hong Kong's sovereignty handover from the United Kingdom to China.

An Empirical Study of Gender Difference

in the Relationship Between Self-Concept and Mathematics Achievement

Enhancing quality of mathematics education has been a priority of the United States for over 40 years (Johnson, 1999). In Goals 2000 announced by the federal government, American students are expected to perform better than their counterparts in any other nations (http://www.ed.gov/legislation/ GOALS2000). In addition, students' belief in their ability to do well in mathematics has been recognized as a key component of mathematical proficiency by the National Council of Teachers of Mathematics (1989). The non-cognitive outcome is required for reporting under the No Child Left Behind (NCLB) Act due to its well-documented linkage to student performance (Public Law 107-110).

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 the general self-concept into academic and nonacademic components. The relationship between academic self-concept and academic achievement has been postulated to be reciprocal and mutually reinforcing (Marsh, Byrne, & Yeung, 1999). Marsh, Hau, and Kong (2002) further noted,

> 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)

To further generalize the reciprocal model, student gender attributes need to be examined under a cross-cultural context. For instance, in Byrnes, Hong, and Xing's (1997) study of Chinese student performance on an American mathematics test, the gender and culture interaction is significant in the research outcome – Despite a large gender difference in the test result from American students, no such a gap was found on the Chinese counterpart. Marsh et al. (2002) concurred that “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).

In comparative education, Hong Kong, a former colony of the United Kingdom, has been identified as a “window city” for mainland China (Lee & Postiglione, 1994). In 1997, Hong Kong terminated its colonial history, and reunited with China. During the historical transition period, Hong Kong participated in the Third International Mathematics and Science Study (TIMSS) in 1995 and a repeat of TIMSS (TIMSS-R) in 1999. The purpose of this investigation is to examine gender difference in the relationship between mathematics achievement and self-concept two years *before* and *after* the sovereignty switch in 1997. 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 modeling. To fill out this void of the research literature, empirical findings from this investigation may not only facilitate generalization of the existing reciprocal model in the cross-culture context, but also articulate student gender attributes in reconfirmation of the relationship between self-concept and academic achievement.

Literature Review

Byrne (1984) stressed motivational properties of academic self-concept, i.e., changes in the self-concept can lead to changes in academic achievement. Reversely, academic self-concept may emerge as a consequence of academic achievement. To verify the “chicken-egg” relationship, Marsh et al. (2002) analyzed a longitudinal database, and reported that “the results of previous research provide general support for a reciprocal effects model” (p. 729).

More specifically, however, comparative studies seemed to disagree on details of the reciprocal relationship (see Wilkins, 2003). For instance, Wilkins, Zembylas, and Travers (2002) analyzed 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 lowest overall beliefs in student self-ability.

In addition, 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). Marsh et al. (2002) made a concerted effort to contrast the reciprocal relationship between Chinese and English middle schools in Hong Kong. But Yip, Tsang, and Cheung (2003) 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" (p. 303).

In history, "Language issues have infected educational practice and policy from the very earliest days of Hong Kong's existence" (Sweeting, 1997, p. 181). To promote the mother-tongue instruction, the colonial government pressed for a greater emphasis on Chinese instruction before the political transition, but faced strong resistance from parents who wanted the instruction in English (Hau, Marsh, Kong, & Poon, 2000). Fok (2001) explained,

> 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. (p. 6)

On the other hand, because elementary schools adopted Chinese as the instructional language, the switch to English in secondary schools often hindered student understanding of the subject contents (Evans, 2000). The situation became even worse when not all subject teachers were able to teach 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).

Two months after the political transition, the new Hong Kong government demanded instruction in Chinese in most secondary schools. To date, no research literature has shed a clear light on this policy implication. One may speculate that students learn more effectively when taught through their mother tongue (see Cummins, 1996; Garcia, 1993), and the improvement of achievement could contribute to enhancement of positive attitude, as suggested by the reciprocal 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” from the late immersion into the second language instruction (e.g., Krashen, 1997; Ogbu, 1992, 1999). Given the involvement of political and cultural contexts, Tao (1994) recollected that “The medium of instruction has been a very controversial issue in Hong Kong for decades” (p. 323).

Marsh’s (1993) analysis in this area concluded that gender differences in specific domains of self-concept were typically consistent with gender stereotypes. Hong Kong is in many ways a mixture of culture between East and West, and “The traditional Chinese culture is male-oriented culture" (Wong, Lam, & Ho, 2002, p. 830). This stereotypic tradition promotes differentiation between boys and girls. On the other hand, Hong Kong was ruled by the 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, Lam, & Ho, 2002, p. 830), which might narrow the gap between the gender attributes.

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). Fortunately, researchers noted that the TIMSS data are appropriate for a thorough study of gender difference (Mullis, Martin, Fierros, Goldberg, & Stemler, 2000). The focus of this study is to use the empirical data to disentangle the reciprocal relationship between self-concept and mathematics achievement at the historical junction of Hong Kong's political transition. Accompanied with the sovereignty handover, the shift of English emphasis is considered in this investigation to contrast the gender difference in the model reconfirmation. Based on the review of current literature, this empirical investigation is designed to facilitate enrichment of the existing knowledge on a profound topic of mathematics education.

Research Questions

Mathematics is a core school subject in Hong Kong. "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, score variance should be partitioned at the student and school levels to facilitate result generalization across the school settings. 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 plausible model of student self-concept and mathematics score under the context of reforming English role in instruction during the political transition?
3. What are 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 8th grade level in 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 the 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” (ch. 6, p. 3).

To enhance comparability between the TIMSS and TIMSS-R results, TIMSS scores have been rescaled at the 8th grade level using the TIMSS-R procedure, and the data are released for comparison at a public website (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 variation of student scores is fairly large across all plausible scores, additional variables should be introduced at the school level to explain the outcome of student performance (Raudenbush & Bryk, 2002). 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).

Statistical Modeling

Development of individual self-concept can be dated back to Socrates' call of "knowing thyself" more than 2000 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-object" (Hamachek, 2000; James, 1890). From the doer's perspective, TIMSS and TIMSS-R gathered "Students' self-perceptions about usually doing well in mathematics" (Supplement 2 of the TIMSS User Guide, section 1, p. 8). Regarding the self as an object, students had a chance to express their feeling of getting bored by mathematics. Besides these indicators of self-concept, students also reported importance of learning English as perceived by self, friends, and parents, and thus, language preference can be assessed by the factor of English push from various aspects. Based on the postulation that "changes in academic self-concept will lead to changes in subsequent academic achievement" (Marsh, Byrne, & Yeung, 1999, p. 155), the reciprocal model has incorporated the language 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, Goldberg, and Stemler (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 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 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 is employed to handle the statistical calculation.

Results

The TIMSS sample contains cases from 3630 male and 3007 female students, and the TIMSS-R has 2585 male and 2534 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 generalized across different schools. The small variance ratio suggests relatively little variation in the educational outcomes among schools within Hong Kong's urban setting (see Table 1).

Insert Table 1 around here

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

Insert Table 2 around here

Table 3 contains parameter estimates for the reciprocal model in Figure 1 using the TIMSS and TIMSS-R databases. In general, results of statistical testing are sensitive to sample sizes, and trivial differences might reach a significant level simply because of a large sample size (see Henson & Smith, 2000). To avoid this problem in statistical reporting, 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 the American Psychological Association to represent the real value differences in dissemination of statistical findings (see Thompson, 1998).

Insert Tables 3 & 4 around here

Discussion

Gender difference is an important factor in self-concept studies across different cultures. Singelis (2000) maintained that a study of self (male or female) has been the primary reason that many disciplines are increasingly embracing cross-cultural perspectives. In designing the international assessment under a multicultural context, proper attention has been given “in 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, the carefully-collected TIMSS data can be employed to examine gender difference in a cross-culture setting.

Since 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 an education system. 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 equivalency of these scores according to the original TIMSS design (Gonzalez & Smith, 1997).

The relatively small variance at the school level seems to agree to similar results from earlier studies of Willms and Raudenbush (1989) and Mortimore, Sammons, Stoll, Lewis and Ecob (1988). In addition, the result appears to be driven by a similar culture emphasis on school education – The Hong Kong society is fairly homogeneous with Chinese accounting for 98% of the population (Poon, 1999), and "Unlike the case in the UK, there is no official classification of social class in Hong Kong" (Lai, 2001, p. 115). In addition, because of Hong Kong's institution of free schooling, equity of education has been enhanced between male and female students in recent years (Post & Pong, 1998).

Furthermore, Wong, Lam, and Ho (2002) observed, "Large-scale studies examining the effects of schooling on gender differences in academic achievements have seldom been conducted in Hong Kong" (Wong, Lam, & Ho, 2002, p. 828). One of the major educational changes accompanied with Hong Kong's political transition is full implementation of a new language policy by the new Hong Kong government (Evans, 2000). The policy has an effect of sharpening the difference between English and Chinese schools. The grey area of using mixed English and Chinese languages was allowed by the old policy before the political transition, and has now been abandoned under the new policy (Evans, 2000). As a result, one may expect to have larger score variances at the school level in TIMSS-R because of implementation of the language separation. Indeed, Table 1 shows a substantial increase in the variance ratio from TIMSS to TIMSS-R. In addition, because the language policy was not designed *for* or *against* a particular gender category, the effect size between male and female students has been consistently small for both TIMSS and TIMSS-R (Table 1).

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 a strong support of 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. Whereas the combined database is approximately twice as large as the sub-data for each gender, the RMR and GFI values remain stable (Table 2), and this cross-examination indicates that the model fitness is unlikely resulted from statistical artifacts pertaining to a particular sample size (Sharma, 1996).

The released TIMSS and TIMSS-R reports already documented gender difference in mathematics achievement in each nation, and no significant gender gap was found from the Hong Kong databases (Beaton, Mullis, Martin, Gonzalez, Kelly, & Smith, 1997; Mullis et al., 2000). That result is reconfirmed in this study by similar factor loadings (λ_{y1}, ..., λ_{y5}) of the equivalent plausible scores across the gender classification (Table 3). Table 3 also shows smaller factor loadings for the TIMSS-R data, and the reduction of the factor loadings could have resulted from existence of additional disturbance variables during the political transition, such as the *sharpened language contrast under the new language policy*, that reduced coherence of the latent variable identification (Table 3). Regarding gender difference in mathematics achievement, Cheng and Seng (2001) 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). (p. 332)

In the released report on gender differences in mathematics, it is interesting to note that the average score has been switched from male scoring higher in TIMSS (Beaton et al., 1997) to female scoring higher in TIMSS-R (Mullis et al., 2000). Whereas 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 model suggests that changes in the achievement pattern will lead to changes in mathematical self-concept. Although the change in the effect size is small, it is interesting to note a switch of the factor loadings (λ_{y6}, λ_{y7}) on the latent factor of self-concept between male and female students (Table 3). For female students, a smaller loading comes from the indicator of "usually doing well in mathematics" ($\lambda_{y6} < \lambda_{y7}$) in TIMSS. In TIMSS-R, the indicator with a smaller loading is switched to student report of "getting bored by mathematics" ($\lambda_{y6} > \lambda_{y7}$). A reverse of this pattern is true for male students (Table 3). Therefore, similar to the reported better mathematics achievement of female students in urban China (Cheng & Seng, 2001), their Hong Kong counterpart also shows a better female scores (Mullis et al., 2000) and more positive self-concept among females after return of the territory sovereignty to China (see $\lambda_{y6} > \lambda_{y7}$ for the TIMSS-R result in Table 3). In line with the improvement of female mathematics education, the effect size also suggests an increase in the reciprocal relation (b) and path coefficients (γ_1, γ_2) for female students in TIMSS-R (Table 4).

In addition, historical trends should be examined to avoid exaggerating the effect solely from the political transition. Before the government handover, Hong Kong was run by the United Kingdom (UK). Rogers, Galloway, Armstrong, and Leo (2000) pointed out a similar trend inside 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 8th grade (Beaton et al., 1997; Mullis et al., 2000). Hence, even though the TIMSS-R data were gathered only two years after Hong Kong's transition, the change of factor loadings between male and female students (Table 3) does not represent a complete reverse of the established trend under the United Kingdom governance. Instead, the change in Table 3 seems to fit the ongoing adjustment of educational outcomes from the cross-cultural context (Cheng & Seng, 2001; Rogers et al., 2000).

Another interesting finding from Table 3 is gender difference on the factor loadings for the *English push* factor. 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 a highly publicized event after Hong Kong's handover (Evans, 2000). In time of crisis, "girls were more highly monitored [by parents] than boys" (Svensson, 2003, p. 300). Consequently, although no gender difference was found from the mother's push indicator in TIMSS (λ_{x1}), the TIMSS-R results showed that the mother's push indicator was the only indicator having gender differences in the factor loading after eruption of the parental protest (Table 3). The different factor loadings seemed to be justified by female receptiveness of the additional parental pressure after Hong Kong's transition.

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 oriented (Evans, 2000), it is no surprise to observe a higher path coefficient toward the achievement factor (see $\gamma_1 > \gamma_2$ in Table 3). Besides the parental push, students were also exposed to a peer influence in Chinese outside of school. In general, "boys were more exposed to deviant peers than girls" (Svensson, 2003, p. 300). The 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 (γ_1 , γ_2), and the corresponding effect size for gender difference is relatively small for all these indexes (see effect sizes for γ_1 , γ_2, and β in Table 3). Whereas 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 that “There were no statistically significant gender differences on the self-concept-achievement relationship” (p. 101).

In summary, this investigation supports Marsh et al.’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 positive reciprocal relation (β) in Figure 1 seems to suggest a mutual enforcement of between mathematics achievement and self-concept, which is in line with the theory developed in the western nations (Marsh, Byrne, & Yeung, 1999). On the other hand, the small β value also characterizes a week linkage from the Hong Kong databases. Since this study is confined to an analysis of two large-scale data separated four years apart, one may speculate if the cultural changes take a longer time to show gender differences in educational outcomes. Under an assumption that the short 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

Allen, N. L., Carlson, J. E., & Zelenak, C. A. (1999). The NAEP technical 1996 technical report.
Washington, DC: National Center for Education Statistics.

Beaton, A., Mullis, I., Martin, M., Gonzalez, E., Kelly, D., & Smith, T. (1996b). Mathematics achievement in the middle school years. Chestnut Hill, MA: Boston College.

Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

Bray, M. (1997). Education and colonial transition: The Hong Kong experience in comparative perspective. Comparative Education, 33 (2), 157-169.

Byrne, B. M. (1984). The general/academic self-concept homological network: A review of construct validation research. Review of Educational Research, 54, 427-456.

Byrnes, J., Hong, L., & Xing, S. (1997). Gender differences on math subtest of the Scholastic Aptitude Test may be culture specific. Educational Studies in Mathematics, 34, 49-66.

Chan, E. (2002). Beyond pedagogy: Language and identity in post-colonial Hong Kong. British Journal of Sociology of Education, 23, 271-285.

Cheng, S. K. & Seng, Q. K. (2001). Gender differences in TIMSS mathematics achievement of four asian nations: A secondary analysis. Studies in Educational Evaluation, 27, 331-340.

Cooley, C. H. (1902). Human nature and the social order. New York: Scribner.

Commins, J. (1996). Bilingual education: What does the research say? In J. Commins (Ed.), Negotiating identifies: Education for empowerment in a diverse society (pp. 97-133). Ontario, CA: California Association for Bilingual Education.

Evans, S. (2000). Hong Kong's new English language policy in education. World Englishes, 19

(2), 185-204.

Evans, S. (2002). The medium of instruction in Hong Kong: Policy and practice in the new English and Chinese streams. Research Papers in Education, 17 (1), 97-120.

Fok, S. C. (2001, October). Education reform in Hong Kong and paradigm shift. Paper presented at the Institute of Asia Pacific Education Development Conference on Education Research, Seoul, Korea.

Garcia, E. E. (1993). Language, culture and education. Review of Research in Education, 19, 51-100.

Gonzalez, E. J., & Smith, T. A. (1997). Users guide for the TIMSS international database. Chestnut Hill, MA: TIMSS International Study Center.

Hamachek, D. (2000). Dynamics of self-understanding and self-knowledge: Acquisition, advantages, and relation to emotional intelligence. Journal of Humanistic Counseling, Education and Development, 38 (4), 230-242.

Hau, K. T., Marsh, H. W., Kong, C. K., & Poon, A. C. S. (2000, April). Late immersion and language of instruction (English vs. Chinese) in Hong Kong high schools: Achievement growth in language and nonlanguage subjects. Paper presented at the annual meeting of American Educational Research Association, New Orleans, LA.

Henson, R. K. & Smith, A. D. (2000). State of the art in statistical significance and effect size reporting: A review of the APA task force report and current trends. Journal of Research and Development in Education, 33 (4), 285-296.

Holliday, W. G. & Holliday, B. W. (2003). Why using international comparative math and science achievement data from TIMSS is not helpful. The Education Forum, 67, 250-

257.

Hong Kong Education Commission (1990). The curriculum and behavioural problems in schools (Report No. 4). Hong Kong: Government Printer.

James, W. (1890). Principles of psychology (Vol. 1). New York: Holt.

Johnson, J. (1999). Speeches and testimony: Statement by the Acting Assistant Secretary for Elementary and Secondary Education. [Online] Available at http://www.ed.gov/ Speeches/04-1999/990428.html [November 5, 2004].

Joreskog, K., & Sorbom, D. (1993). LISREL8 user's reference guide. Chicago: SSI.

Kifer, E. A. (2002). Students' attitudes and perceptions. In D. F. Robataille & A. E. Beaton (Eds.), Secondary Analysis of the TIMSS Results: A Synthesis of Current Research. Boston, MA: Kluwer Academic Publishers.

Krashen, S. (1997). Why bilingual education? ERIC Digest. Charleston, WV: ERIC Clearinghouse on Rural Education and Small Schools.

Lai, M. (2001). Hong Kong students' attitudes towards Cantonese, Putonghua and English after the change of sovereignty. Journal of Multilingual and Multicultural Development, 22 (2), 112-133.

Lee, W. O. & Postiglione, G. A. (1994). The "Window Effect" on education and development in South China. Chinese University Education Journal, 21, 111-121.

Ma, X. & Kishor, N. (1997). Attitude toward self, social factors, and achieveent in matheatics: A

meta-analytic review. Educational Psychology Review, 9 (2), 89-120.

Marsh, H. W. (1993). The multidimensional structure of academic self-concept: Invariance over gender and age. American Educational Research Journal, 30, 841-860.

Marsh, H. W., Byrne, B. M., & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement: Reanalysis of a pioneering study and revised recommendations. Educational Psycholigist, 34 (3) 155-167.

Marsh, H. W., Hau, K. T., & Kong, C. K. (2000). Late immersion and language of instruction in Hong Kong high schools: Achievement growth in language and nonlanguage subjects. Harvard Educational Review, 70 (3), 302-346.

Marsh, H. W., Hau, K. T., & Kong, C. K. (2002). Multilevel causal ording of academic self-concept and achievement: Influence of language of instruction (English compared with Chinese) for Hong Kong students. American Educational Research Journal, 39, 727-763.

Martin, M., Mullis, I., Gregory, K., Hoyle, C., & Shen, C. (2001). Effective schools in science and mathematics. Chestnut Hill, MA: Boston College.

Mead, G. H. (1934). Mind, self, and society. Chicago: University of Chicago Press.

Morris, P. & Chan, K. K. (1997). The Hong Kong school curriculum and political transition: Politicisation, contextualisation and symbolic action. Comparative Education, 33 (2), 247-264.

Mortimore, P., Sammons, P., Stoll, L., Lewis., D., & Ecob, R (1988). The effects of school membership on pupils' educational outcomes. Research Papers in Education, 3, 3-26.

Mullis, I., Martin, M., Gonzalez, E., Gregory, K., Garden, R., O'Connor, K., Chrostowski, S., & Smith, T. (2000). TIMSS 1999 international mathematics report: Findings from IEA's repeat of the Third International Mathematics and Science Study at the eighth grade. Chestnut Hill, MA: Boston College.

Mullis, I. V. S., Martin, M. O., Fierros, E., G., Goldberg, A. L., & Stemler, S. E. (2000). Gender differences in achievement: IEA's Third International Mathematics and Science Study (TIMSS). Chestnut Hill, MA: Boston College.

National Council of Teachers of Mathematics. (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: National Council of Teachers of Mathematics.

Ogbu, J. U. (1992). Understanding cultural diversity and learning. Educational Researcher, 21, 5-14.

Ogbu, J. U. (1999). Beyond language: Ebonics, proper English, and identity in a Black-American speech community. American Educational Research Journal, 36, 147-184.

Poon, A. (1999). Chinese medium instruction policy and its impact on English learning in post-1997 Hong Kong. International Journal of Bilingual Education and Bilingualism, 2, 131-146.

Post, D. & Pong, S. (1998). The waning effect of sibship composition on school attainment in Hong Kong. Comparative Education Review, 42, 99-117.

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

Rogers, C. G., Galloway, D., Armstrong, D., & Leo, E. (2000). Gender differences in motivational style. European Education, 32, 79-98.

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

Sharma, S. (1996). Applied multivariate techniques. New York: John Wiley & Sons.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (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.

Svensson, R. (2003). The impact of parental monitoring and peer deviance. Youth & Society, 34, 300-329.

Sweeting, A. (1997). Education policy and the 1997 factor: The art of the possible interacting with the dismal science. Comparative Education, 33, 171-185.

Tao, P. K. (1994). Comprehension of non-technical words in science: The case of students using a ‘foreign’ language as the medium of instruction. Research in Science Education, 24, 322-330.

Thompson, B. (1998). Statistical significance and effect size reporting: Portrait of a possible future. Research in the Schools, 5 (2), 33-38.

Wang, J. (2001, August-September). TIMSS primary and middle school data: Some technical concerns. Educational Researcher, 17-21.

Wang, J. (2004, April). Self-concept and mathematics achievement: Modeling the relationship under the language pressure in Hong Kong. Paper presented at the annual meeting of the American Educational Research Association (AERA), San Diego, CA.

Wang, J., Oliver, S., & Garcia, A. (2004, April). An empirical study of relationships between student self-concept and science achievement in Hong Kong. Paper accepted for presentation at the annual meeting of the American Educational Research Association(AERA), San Diego, CA.

Wigfield, A., Eccles, J. S., MacIver, D., Reuman, D. A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain specific self-perceptions and general self-esteem across the transition to junior high school. Developmental

Psychology, 27, 552-565.

Wilgenbusch, T. & Merrell, K. W. (1999). Gender differences in self-concept among children and adolescents: A meta-analysis of multidimensional studies. School Psychology Quarterly, 14, 101-120.

Wilkins, J. L. M. (2003, April). The relationship between mathematics and science self-concept and related achievement: An international investigation. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Wilkins, J. L. M., Zembylas, M., & Travers, K. J. (2002). Investigating correlates of mathematics and science literacy in the final year of secondary school. In D. F. Robataille & A. E. Beaton (Eds.), Secondary Analysis of the TIMSS Results: A Synthesis of Current Research. Boston, MA: Kluwer Academic Publishers.

Willms, J. & Raudenbush, S. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. Journal of Educational Measurement, 26, 209-232.

Wong, K.C., Lam, Y. R., & Ho, L. M. (2002). The effects of schooling on gender differences. British Educational Research Journal, 28, 827-843.

Yip, D. Y., Tsang, W. K., & Cheung, S. P. (2003). Evaluation of the effects of medium of instruction on the science learning of Hong Kong secondary students: Performance on the science achievement test. Bilingual Research Journal, 27, 295-331.

Younger, M., Warrington, M., & Williams, J. (1999). The gender gap and classroom interactions: Reality and rhetoric? British Journal of Sociology of Education, 20, 325-341.

Table 1

Ratio of variances in mathematics plausible scores (PS) distributed at school and student levels[1]

Project	Gender	PS1	PS2	PS3	PS4	PS5	Mean
TIMSS	Male	.050	.049	.051	.049	.047	.049
	Female	.047	.047	.049	.046	.047	.047
TIMSS-R	male	.085	.085	.089	.087	.082	.085
	Female	.083	.090	.085	.087	.086	.086

[1] The ratio is computed by dividing the partitioned score variance at the school level over the corresponding variances at the student level.

Table 2

Model fitting indices for the TIMSS and TIMSS-R databases

	TIMSS Samples			TIMSS-R Samples		
	Total	Male	Female	Total	Male	Female
RMR	0.10	0.09	0.10	0.10	0.09	0.11
GFI	0.98	0.98	0.98	0.98	0.98	0.97

Table 3

Parameter estimates for the reciprocal model using the TIMSS and TIMSS-R databases

Estimates	TIMSS		Effect	TIMSS-R		Effect
	Male	Female	Size	Male	Female	Size
Factor Loadings						
λ_{y1}	.96	.96	.00	.93	.92	.01
λ_{y2}	.96	.95	.01	.93	.92	.01
λ_{y3}	.96	.96	.00	.93	.92	.01
λ_{y4}	.96	.95	.01	.93	.92	.01
λ_{y5}	.96	.96	.00	.94	.92	.02
λ_{y6}	.62	.57	.05	.51	.66	–.15
λ_{y7}	.56	.63	–.07	.78	.64	.14
λ_{x1}	.64	.64	.00	.59	.64	–.05
λ_{x2}	.57	.66	–.09	.63	.63	.00
λ_{x3}	.74	.78	–.04	.83	.83	.00
Reciprocal Relation						
β	.11	.05	.06	.07	.08	–.01
Path Coefficients						
γ_1	.36	.23	.13	.34	.30	.04
γ_2	.28	.21	.07	.22	.28	–.06

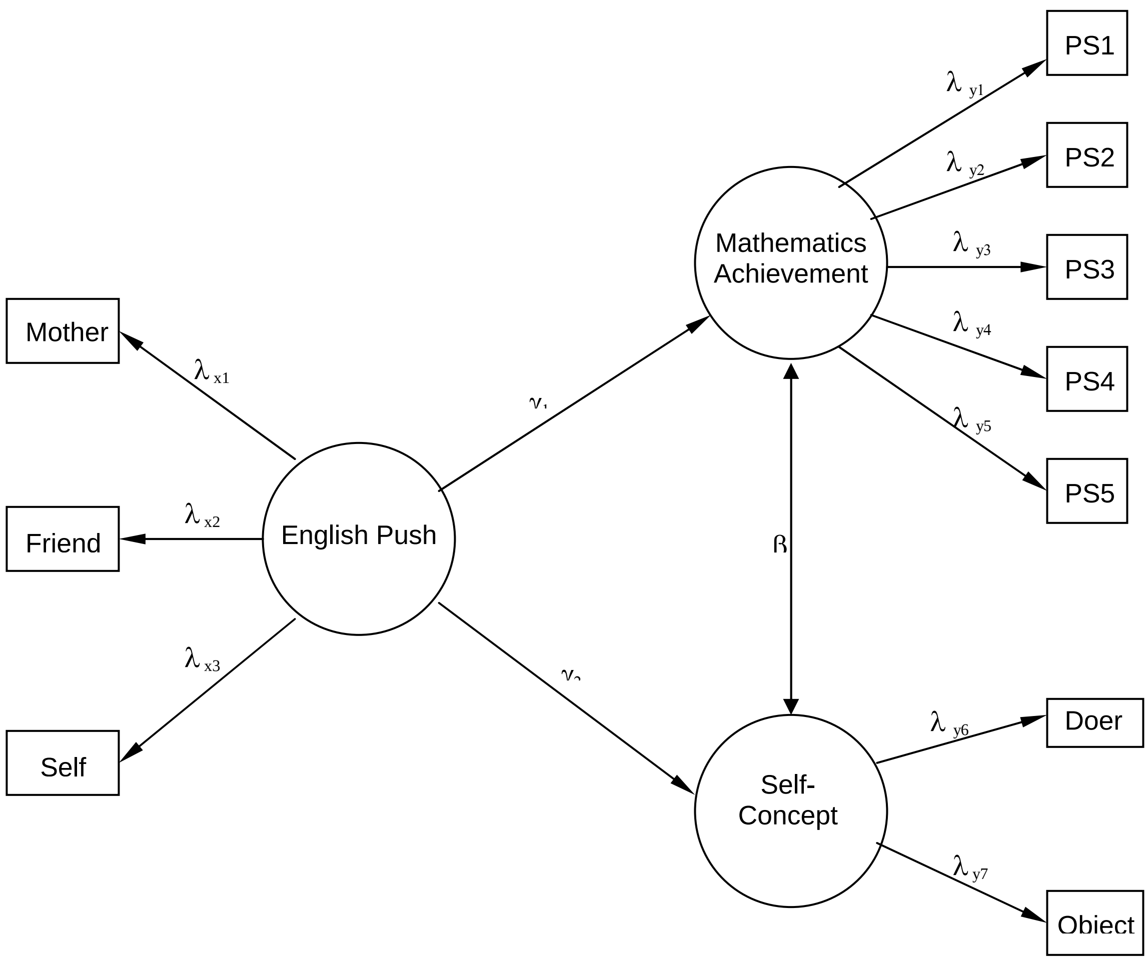
Table 4

Effect sizes of the structural relations from the period of Hong Kong's political transition

Gender	β	γ_1	γ_2
Male	-.04	-.02	-.06
Female	.03	.07	.07

Figure 1

A Structural Model of Self-Concept and Mathematics Achievement from TIMSS



Notes:

1. PS1, … PS5 are plausible mathematics scores imputed from TIMSS or TIMSS-R projects under a three-
2. The "doer" and "object" indicators are based on student responses on whether they can do well in mathematics and felt bored about mathematics.
3. Indicators for the "English push" factor are based on student responses on importance of learning English perceived by self, friends, and parents.