Relations Between Academic Self-concept and Achievement in Mathematics and Language: Cross-Cultural Generalisability

Herbert W. Marsh
Self-concept Enhancement and Learning Facilitation Research Centre
University of Western Sydney, Australia

Kit-Tai Hau
The Chinese University of Hong Kong, Hong Kong

Paper presented at NZARE AARE, Auckland, New Zealand, November 2003
HAU03756

Relations Between Academic Self-concept and Achievement in Mathematics and Language: Cross-Cultural Generalisability

Herbert W. Marsh
Self-concept Enhancement and Learning Facilitation Research Centre
University of Western Sydney, Australia

Kit-Tai Hau
The Chinese University of Hong Kong, Hong Kong

Despite students' achievements in mathematics and language being positively related, the corresponding self-concepts are nearly uncorrelated. The universality of this paradoxical pattern of relations was examined in a large cross-cultural study of nationally representative samples of 15-year-olds from 26 countries with a total of more than 50,000 students. In this Internal/External Frame (I/E) Model, it was also postulated that math achievement had positive effects on math self-concept, but negative effects on verbal self-concept, and vice versa for language achievement on language and math self-concepts. Multigroup structural equation models demonstrated good support for the invariance of results across 26 countries participating in this PISA project.

Self-concept has been an important educational outcome indicator (high self-concept itself is good) as well as a desirable mediating variable leading to other positive outcomes. For example, educational policy statements throughout the world list self-concept enhancement as a central goal of education and an important vehicle for addressing social inequities experienced by disadvantaged groups. The purpose of the present investigation is to examine the cross-cultural generalizability of relations between math and verbal self-concepts with math and verbal achievements in a large, nationally representative samples of 15-year-olds from 26 countries. This is important because previous self-concept research and support for the construct validity of major instruments and the main theoretical models has been based largely on responses by students from Western countries – particularly English-speaking students in the United States, Australia, and Canada. In this study, the internal/external frame of reference (I/E) model was developed to explain what initially seemed to be paradoxical patterns of relations between Math and Verbal self-concepts and corresponding areas of achievement.

Relation of Self-Concept to Achievement and Internal/External Frame of Reference

In the Shavelson Huber and Stanton (1976) model, self-concept is posited to be a multidimensional, hierarchical construct. Global self-concept, at the apex of the hierarchy, is divided into nonacademic (e.g., social, physical, emotional) and academic components. Of particular relevance to the present investigation, academic self-concept is divided into self-concepts in particular content areas such as math and verbal self-concepts. Support for the construct validity of self-concept interpretations and its multidimensionality requires that: (a) academic achievement is more highly correlated with academic components of self-concept than with global and nonacademic components of self-concept; (b) academic achievement in particular domains is more highly correlated with academic self-concepts in the matching domain (e.g., math achievement and math self-concept) than self-concepts in nonmatching domains (e.g., math achievement and general or verbal self-concept).

Internal/External Frame Of Reference (I/E) Model

In Shavelson et al.'s (1976) model, the assumption of a strong higher-order academic self-concept implies that there is a substantial correlation between verbal and math self-concepts. This prediction also follows from the typically large correlation between math and verbal academic achievements (typically .5 to .8, depending on how achievement is measured). Early research, however, demonstrated that math and verbal self-concepts were much more differentiated than the corresponding achievement scores (Marsh, 1986). In contrast to the expectation of high correlations between math and verbal self-concepts, math and verbal self-concepts were nearly uncorrelated. Furthermore, this near-zero correlation was consistent across different measures of the math and verbal self-concepts and a diversity of settings (Marsh & Craven, 1997; Marsh & Yeung, 1998). Hence, it seems that individuals with good mathematics skills also tend to have good verbal skills and vice-versa, but people think of themselves as either "math" persons or "verbal" persons – but not both.

The I/E model (for further discussion, see Marsh, 1986, 1990a, 1993; Marsh, Byrne, & Shavelson, 1988; Marsh & Yeung, 2001) was initially developed to explain why math and verbal self-concepts are almost uncorrelated even though corresponding areas of academic achievement are substantially correlated. According to the I/E model, academic self-concept in a particular domain (e.g., math or verbal self-concepts) is formed in relation to two comparison processes or frames of reference.
The first ***external*** (normative) reference is the typical social comparison process in which students compare their self-perceived performances in a particular school subject with the perceived performances of other students in the same school subject and other external standards of actual achievement levels (e.g., normative comparisons, school grades, class rankings, etc.). If they perceive themselves to be able in relation to other students and objective indicators of achievement, then they should have a high academic self-concept in that school subject. In the second ***internal*** (ipsative-like) reference, students compare their own performance in one particular school subject with their own performances in other school subjects. If, for example, mathematics is their best school subject, then they should have a positive math self-concept that is higher than their verbal self-concept. Similarly, according to this internal comparison process, students may have a favorable math self-concept if math is their best subject even if they are not particularly good at math relative to other students and external standards. It is this internal comparison process that represents an extension to traditional social comparison theory.

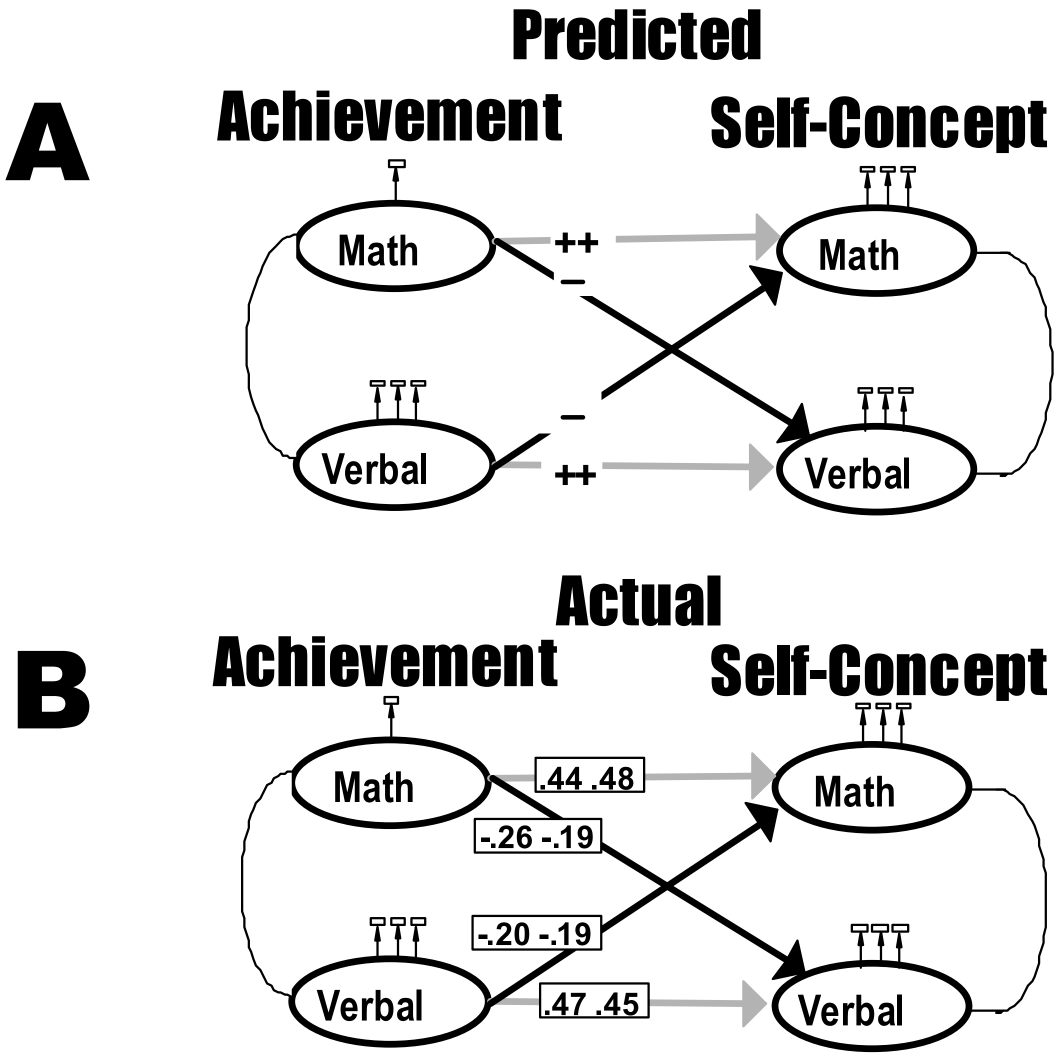
Depending upon how these two processes are weighted in the formation of self-concept, these students may have an average or even above-average math self-concept even though they have below-average math skills. The I/E model also predicts that these students would have better math self-concepts than other students who did equally poorly at mathematics but who did better in all other school subjects (i.e., math was their worst subject). Similarly, a student who is very bright in all school subjects may have an average or even below-average math self-concept if the student perceived mathematics to be his or her worst subject.

The external comparison process should result in substantial positive correlations between math and verbal self-concepts because math and verbal achievements are substantially positively correlated. However, the ipsative, internal comparison process should result in a negative correlation between math and verbal self-concepts because the average correlation among ipsative scores is necessarily negative (i.e., an increase in any one score must result in the counterbalancing decrease in average of the remaining scores if they are ipsative). Both these processes, however, affect self-concept responses. Hence, the joint operation of these processes, depending on the relative weight given to internal and external comparisons, is consistent with the near-zero correlation between math and verbal self-concepts that led to the development of the I/E model. It is, however, important to emphasize that support for the I/E model does not require the correlation between math and verbal self-concepts to be zero, but only that it be substantially less than the typically substantial positive correlation between math and verbal achievement.

Domain specificity is a critical feature of the I/E model. However, stronger tests of the I/E model are possible when math and verbal achievements are related to math and verbal self-concepts (see Figure 1A). The external comparison process predicts that good math skills lead to higher math self-concepts and that good verbal skills lead to higher verbal self-concepts. According to the internal comparison process, however, good math skills should lead to *lower* verbal self-concepts (once the positive effects of good verbal skills are controlled). The better I am at mathematics, the poorer I am at verbal subjects (relative to my good math skills). Similarly, better verbal skills should lead to *lower* math self-concept (once the positive effects of good math skills are controlled). In models used to test this prediction (Figure 1A), the horizontal paths leading from math achievement to math self-concept and from verbal achievement to verbal self-concept (the gray horizontal lines in Figure 1A) are predicted to be substantially positive (indicated by "++" in Figure 1A). However, the cross paths leading from math achievement to verbal self-concept and from verbal achievement to math self-concept (the dark lines in Figure 1A) are predicted to be negative. Although consistent with the I/E model, it is these negative cross paths – the negative effects of verbal achievement on math self-concept and of math achievement on verbal self-concept – that initially appeared to be paradoxical and provide the critical test of the I/E model.

The critical prediction for the I/E model, however, is in terms of the path coefficients. In other words, it is the effect of math achievement on verbal self-concept AFTER controlling for the effect of verbal achievement. Hence, once we control for the positive effect of verbal achievement on verbal self-concept, then the unique component of math achievement is negatively associated with verbal self-concept. Thus, the size of the negative effect of math achievement on verbal self-concept should be a function of the discrepancy between the math and verbal achievement scores. Hence, the operative construct is a residual score that is conceptually like the difference score (without some of the statistical problems associated with raw difference scores). Thus, a B-average in mathematics may induce an average or even below-average mathematics self-concept for the student who earns A's in most other school subjects, but may lead to an above-average math self-concept for the student who earns C's in other subjects. In the language of path analysis, it is the direct effect of math achievement on verbal self-concept that is predicted to be negative – not the correlation between math achievement and verbal self-concept.

Figure 1. **Predicted (A) and actual (B) results based on the I/E model.**



In A the horizontal (positive) paths are predicted to be substantial and positive ("++"), whereas the cross (negative) paths are predicted to be smaller and negative ("–"). In B, the actual results based on the total group analysis (Model TG1 in Table 1) and the multiple group analysis (Model MG5 in Table 1) are consistent with predictions.

In a review of 13 studies that considered students of different ages and different academic achievement indicators, Marsh (1986) reported that: (a) correlations between indicators of verbal and math achievement were substantial (.42 to .94); (b) correlations between measures of Verbal and Math self-concepts were much smaller (-.10 to +.19); (c) path coefficients from verbal achievement to Verbal self-concept, and from math achievement to Math self-concept were all significantly positive; (d) path coefficients from math achievement to verbal self-concept, and from verbal achievement to math self-concept, were significantly *negative*.

This pattern of results consistent with the I/E model was subsequently replicated for responses to each of three different self-concept instruments by Canadian high school students (Marsh et al., 1988), for the nationally representative sample of US high school students in the High School and Beyond Study (Marsh, 1989), for the nationally representative sample of US high school students in the National Longitudinal Study (Marsh, 1994b), and for academically gifted students in the US (Plucker & Stocking, 2001; Williams & Montgomery 1995) and China (Lee, Yeung, Low & Jin, 2000). Although much of this support is based on responses by students from United States, Canada, and Australia where the native language is English, there is also some support for the cross-cultural or cross-nationality generalizability of these results where Verbal self-concept is based on a native language other than English (e.g., Norway: Skaalvik & Rankin, 1995; Skaalvik & Valas, 2001; China: Dai, 2001; 2002; Kong, 2000; Lee, Yeung, Low & Jin, 2000; Marsh, Kong & Hau, 2001; Yeung & Lee, 1999; Yeung & Lau, 1998; Germany: Marsh & Köller, in press; Moeller & Köller 2001; United Arab Emirates: Abu-Hilal, 2002; Abu-Hilal & Bahri, 2000). Whereas these studies have found support for the I/E model in different countries and cultural groups, each was based on results from a single country and a methodology (e.g., achievement indicators, instrumentation, translation, selection and representativeness of the sample, statistical analysis) that was at least partially idiosyncratic to the particular study. In their critique of research in this area, Marsh and Yeung (2001) noted the need to systematically pursue cross-national and cross-cultural comparisons of this research in order to evaluate more fully the generalisability of support based in the I/E model.

Methods

Subjects

The study was based on a nationally representative sample of 15-year olds collected in 32 countries in the year 2000 (see OECD, 2001a, 2001b, for a description of the database and variables) in the Program of Student Assessment (PISA) project coordinated by Organisation For Economic Co-Operation And Development (OECD). The PISA database was collected in response to the need for internationally comparable evidence of student performance and related competencies within a common framework that is internationally agreed upon. Selection of the measures was made on the basis of advice from substantive and statistical expert panels and results from extensive pilot studies. Substantial efforts and resources were devoted to achieving cultural and linguistic breadth in the assessment materials, stringent quality-assurance mechanisms were applied in the translation of materials into different languages, and data were collected under independently supervised test conditions. Paper-and-pencil assessments consisted of a combination of multiple choice items and written responses.

For the purpose of the present investigation, we analyzed only students who had both mathematics and reading achievement test scores and who completed the math and verbal self-concept items. Measures included three measures of reading achievement, a single measure of mathematics, three math self-concept items and three verbal self-concept items. The self-concept items were from the highly regarded Self Description Questionnaire II (Byrne, 1996; also see Marsh, 1990b; 1993). Although 97,384 students completed math and reading assessments, only 59,332 also completed any of the self-concept items and 55,577 had complete data for all ten variables considered here (i.e., *N* after listwise deletion for missing data, the basis of the present investigation). As recommended in the database documentation (OECD, 2001a, 2001b), all analyses of the PISA data should be weighted to obtain unbiased estimates of population parameters. For purposes of the present investigation, the effective sample size for each country was set equal to the number of cases for that country prior to weighting so that the weighted sample size was the same as the unweighted sample size (i.e., the average weight across all cases was 1.0).

Analysis

The data were analyzed with structural equation models (SEMs) conducted with LISREL 8 (Joreskög and Sorböm) using maximum likelihood estimation (for further discussion of SEM, see Bollen, 1989; Byrne, 1998; Joreskög and Sorböm). Following Marsh, Balla, and Hau (1996), and Marsh, Balla, and McDonald (1988) we emphasize the Tucker-Lewis index (TLI), the relative noncentrality index (RNI), and root mean square error of approximation (RMSEA) to evaluate goodness of fit, but also present the χ^2 test statistic and an evaluation of parameter estimates. Whereas tests of statistical significance and indices of fit aid in the evaluation of the fit of a model, there is ultimately a degree of subjectivity and professional judgment in the selection of a ”best” model (Marsh, Balla & McDonald, 1988).

When there are parallel data from more than one group – the 26 countries in this study -- it is possible to test the invariance of the solution by requiring any one, any set, or all parameter estimates to be the same in two or more groups. Byrne (in press) argued that this analysis is particularly appropriate for making cross-cultural comparisons. In applying this approach (e.g., Byrne, 1998; Marsh, 1994a), there is a well developed methodology in which the goodness of fit of alternative models are compared, including the least restrictive model that does not require any of the parameter estimates to be the same in different groups and the most restrictive model that requires all parameter estimates to be the same in the different groups. Typically, the minimal condition for "factorial invariance" is the equivalence of all factor loadings in the multiple groups and this is one of the first tests of invariance in the sequence. There is no clear consensus in recommendations about the ordering of subsequent invariance constraints (e.g., Bentler, 1988; Bollen, 1989; Byrne, 1998; Joreskög and Sorböm, 1993), although Bentler (1988; Byrne, 2001) noted that the equality of parameters associated with measurement errors is typically the least important hypothesis to test and is unlikely to be met in most applications. In the present investigation, the four path coefficients used to test the I/E model (Figure 1) are of critical importance and are the focus of specific models designed to evaluate the invariance of these parameters across the 26 countries.

Coefficient alpha estimates of reliability, computed separately for each country, were consistently high across the 26 groups for reading achievement (M = .86, SD = .03), math self-concept (M = .88, SD = .02), and verbal self-concept (M = .74, SD = .07). Inspection of the item-total correlations for each indicator (not shown) indicated that the one negatively worded (reverse scored) item in the verbal self-concept scale contributed less positively to reliability than the other two positively worded verbal self-concept items (all math self-concept items were positively worded) and that this pattern of results was consistent across the 26 countries. In the PISA database, mathematics achievement was represented by a single score. Therefore. for purposes of structural equation models in the present investigation, we set its reliability at .90 (see discussion by Joreskög and Sorböm).

Results and Discussion

A SEM analysis with the total group of 55,582 participants (Tables 1 and 2) showed that this total group model (TG1) was well defined and the goodness of fit (TLI = .97; see Table 1) was very good. Most important, however, were tests of the four path coefficients that were central to the evaluation of support for the I/E model. As predicted, the two horizontal paths relating math achievement to math self-concept (.44) and relating reading achievement to verbal self-concept (.47) were substantial and positive, whereas the two cross paths leading from Reading achievement to Math self-concept (-.20) and Mathematics achievement to verbal self-concept (-.26) were negative. Also of relevance is the observation that the correlation between Math and Verbal achievement factors (r = .78) was very large, whereas the correlation between Math and Verbal self-concept factors (r = .10) was substantially lower. Hence, results based on the total sample clearly support the main predictions for the I/E model (see Figure 1).

Generalizability Over Countries with Tests of Invariance
The overarching question is whether the I/E model is general enough to be replicable in the 26 countries being examined. To pursue this question, we conducted multigroup CFAs and SEMs in which we constrained different parameters to be invariant across the 26 groups (Table 2). We began with a set of CFA models to evaluate the invariance of the measurement component of the model and then focused specifically on structural equation models to evaluate the I/E model.

No invariance constraints were imposed on the parameters in the baseline multiple-group model (MG1). The fit indexes for this model (e.g., TLI = .97) were very good. In the first test of invariance (model MG2), factor loadings were constrained to be equal across the 26 groups. Again, the fit indexes were very good and differed little from those based on the totally non-invariant solution (MG1). This supports the appropriateness of the measures across the 26 groups and satisfies the minimum requirement for factorial invariance. In each of the subsequent CFA models (MG3 – MG6 in Table 2), the invariance of the factor loadings was imposed in combination with the invariance of additional sets of parameters – factor variances, factor covariances, and uniquenesses. Although the imposition of these added invariance constraints resulted in small decrements in fit, even the highly restrictive model MG6 of total invariance (i.e., requiring every parameter to be the same in all 26 groups) provided a good fit to the data that differed only slightly from Model MG1 that had no invariance constraints. These results support the cross-cultural generalizability of the measures and the relations among them across these 26 countries.

***Table 1*: Parameter Estimates for Total Group Solution and Multiple Group Solution**

	Total Group Solution					Multiple Group Solution			
	MAch	VAch	MSC	VSC	Uniq	MAch	Vach	MSC	VSC
Factor Loadings									
MAch	.95	.00	.00	.00	.10	.94	.00	.00	.00
VAch1	.00	.85	.00	.00	.28	.00	.83	.00	.00
VAch2	.00	.89	.00	.00	.22	.00	.87	.00	.00
VAch3	.00	.78	.00	.00	.39	.00	.77	.00	.00
MSC1	.00	.00	.84	.00	.29	.00	.00	.84	.00
MSC2	.00	.00	.85	.00	.27	.00	.00	.85	.00
MSC3	.00	.00	.83	.00	.30	.00	.00	.83	.00
VSC1	.00	.00	.00	.55	.70	.00	.00	.00	.60
VSC2	.00	.00	.00	.72	.48	.00	.00	.00	.71
VSC3	.00	.00	.00	.83	.31	.00	.00	.00	.81
Path Coefficients									
MAch	.00	.00	.00	.00		.00	.00	.00	.00
VAch	.00	.00	.00	.00		.00	.00	.00	.00
MSC	.44	-.20	.00	.00		.48	-.19	.00	.00
VSC	-.26	.47	.00	.00		-.19	.45	.00	.00
Variance/Covariances									
MAch	1.00					1.00			
VAch	.78	1.00				.76	1.00		
MSC	.00	.00	.90			.00	.00	.87	
VSC	.00	.00	.11	.90		.00	.00	.04	.89

Note. MAch =Math Achievement, VAch = Verbal Achievement, MSC = Math Self-concept, VSC = Verbal Self-concept, Uniq = uniqueness. All parameter estimates are present in completely standardized form. The total group solution is based on Model TG1 (Table 2) and the multiple group solution is based on MG3 (with invariant factor loadings, path coefficients, and factor variance/covariances, but freely estimated uniquenesses for each of the 26 countries).

***Table 2*: Goodness of Fit for I/E Model fit to the Total Group and Multiple (country) Groups**

MODEL	CHISQ	DF	RNI	TLI	RMSEA	Model Description
Total Sample						
TG1	5026.06	30	.98	.97	.05	Full I/E Model
Multiple Group CFA						
MG1	5784.36	780	.98	.97	.05	CFA INV=none; Free= FL, FV, FC, Uniq
MG2	7650.34	930	.97	.97	.06	CFA INV=FL; Free = FV, FC, Uniq.
MG3	9846.64	1030	.97	.96	.06	CFA INV=FL, FV; Free = FC, Uniq.
MG4	9070.34	1005	.97	.96	.06	CFA INV=FL, FC; Free = FV, Uniq.
MG5	12515.47	1180	.96	.96	.07	CFA INV=FL, FC, FV; Free =Uniq
MG6	18513.56	1405	.93	.95	.08	CFA INV=FL, FC, FV, uniq (total invariance)
Multiple Group SEM						
MG7	9497.48	1030	.97	.96	.06	SEM INV=FL, PC; Free = FV, FC, uniq.
MG8	8078.61	980	.97	.97	.06	SEM INV=FL, PC-; Free = FV, FC, uniq, PC+
MG9	8273.18	980	.97	.97	.06	SEM INV=FL, PC+; Free = FV, FC, uniq, PC-
MG10	10577.99	1080	.96	.96	.06	SEM INV=FL, FV, FC; Free = PC, uniq
MG11	10520.63	1080	.96	.96	.06	SEM INV=FL, FC, PC; Free = FV, uniq
MG12	11445.96	1130	.96	.96	.07	SEM INV=FL, FV, PC; Free = FC, uniq
MG13	11234.89	1130	.96	.96	.06	SEM INV=FL, FV, FC, PC+; Free = PC-, uniq
MG14	11153.97	1130	.96	.96	.06	SEM INV=FL, FV, FC, PC-; Free = PC+, uniq
MG15	12515.47	1180	.96	.96	.07	SEM INV=FL, FV, FC, PC; Free = uniq

Note. RNI = relative noncentrality index, TLI = Tucker-Lewis index, RMSEA = root mean square error of approximation, DF = degrees of freedom, CFA = Confirmatory Factor Analysis, SEM = Structural Equation Model, FL = Factor loading, FC = factor covariances, FV = Factor Variances, PC = Path Coefficient, PC+ = Horizontal Path Coefficients predicted to be positive (see Figure 1), PC- = Cross Path Coefficients predicted to be negative (see Figure 1), Uniq = uniqueness. In Model TG1 (see parameter estimates in Table 1) the I/E model was fit to the total group, whereas for Models MG1-MG13 the I/E model was fit separately for each of the 26 groups representing different countries. For Models MG2-MG13, some combination of parameters is required to be invariant across the 26 groups (countries).

Models MG7-MG15 focus specifically on the structural component of the model – the path coefficients that are critical to tests of predictions based on the I/E model (see Figure 1). In model MG7, the path coefficients and factor loadings were required to be the same in each of the 26 groups. Although there was a very small decrement in fit (TLI = .96) relative to the model with only factor loadings invariant (MG1), the fit was still very good. These tests of the invariance of the four path coefficients provided a global test of the invariance of the two path coefficients predicted to be positive and the two path coefficients predicted to be negative. In Model MG8 the horizontal (positive) path coefficients were freely estimated in each group whereas the cross (negative) path coefficients were required to be the same in all 26 groups. In Model MG9, the negative path coefficients were freely estimated and the positive paths were invariant across groups. In both models the goodness of fit improved a small amount (both TLIs are .97), but the differences were small. These results demonstrated that the magnitudes – as well as the direction – of the path coefficients were consistent across the 26 different countries.

In Models MG10-MG12, we evaluated the effects on goodness of fit associated with invariance constraints on factor variances and factor covariances in the I/E model. Whereas these additional invariance constraints produced some decrement in fit, even Model MG15 (that required that all four path coefficients, all four factor variances, and both factor covariances were the same across the 26 countries) provided a good fit to the data.

In summary, even the extremely demanding model with complete invariance of all parameters provided a good fit to the data. Because no one of these multiple group models stood out as clearly the “best” model, we evaluated parameter estimates based on several of these models.

Generalizability of Parameter Estimates

Further analyses were conducted to evaluate further the cross-cultural generalizability of the parameter estimates based on model MG5 (Table 1). Because factor loadings, path coefficients, and factor variances and covariances are invariant (the same) across the 26 groups, it is only necessary to present one set of parameter estimates (rather than separate sets of parameter estimates for each of the 26 groups). Because uniqueness terms were not held invariant across groups in this model, the 26 separate sets of uniqueness terms are not presented in order to conserve space (but see earlier discussion of reliability estimates; also see Table 3). Parameter estimates for this highly restrictive multigroup model MG5 were nearly the same as those based on the total group model TG1. Of particular importance, are the cross (negative) paths leading from Reading achievement to Math self-concept (-.19) and from Mathematics achievement to verbal self-concept (-.19). In addition to providing global support for the I/E model, the invariance of these parameter estimates provides remarkably strong support for the cross-cultural generalizability of predictions based on the I/E model.

Of critical importance to the present investigation, were the four path coefficients relating the two achievement test scores to the corresponding self-concept measures. Although the fit was good for models that required these path coefficients to be the same across the 26 countries, this highly restrictive model produced a small decrement in fit. In order to evaluate the extent of variation in different countries, path coefficients from Model MG10 (which allowed the path coefficients to be estimated separately in each country) are presented for all 26 countries in Table 3.

Horizontal (positive) paths from math achievement to math self-concept and from verbal achievement to verbal self-concept were predicted to be substantial and positive. All 52 of these path coefficients were statistically significant and positive. The means of the two sets of path coefficients were .51 (SD = .18) and .47 (SD = .12) respectively. Cross (negative) paths from math achievement to verbal self-concept and from verbal achievement to math self-concept were predicted to be negative and less substantial. Across these 52 path coefficients, one was small but significantly positive (.12), 7 were non-significant, and the remaining 44 were significantly negative. The means of the two sets of path coefficients were -.22 (SD = .16) and -.21 (SD = .14) respectively. These results for horizontal and cross paths were similar to those based on the total sample and those based on Model MG15 in which path coefficients were required to be the same across the 26 groups (see Table 1). In all 26 countries, the absolute sizes of the cross (negative) paths were consistently much smaller than the systematically larger horizontal (positive) paths.

***Table 3*: Reliability Estimates and Coefficients For Each Country: Model MG2 (Table 1; also see Figure 1)**

		Reliability			Factor Corr		Path Coefficients			
							From	From	From	From
		VAcH	MSC	VSC	MAch	VSC	MAch	VAch	MAch	VAch
Country	N	α	α	α	Vach	MSC	to MSC	to MSC	to VSC	to VSC
Total	**55582**	**.87**	**.88**	**.74**	**.78***	**.10***	**.48***	**-.19***	**-.19***	**.45***
1 Australia	2642	.87	.86	.78	.77*	.08*	.41*	-.16*	-.19*	.39*
2 Austria	2380	.87	.88	.81	.76*	-.07*	.47*	-.25*	-.26*	.56*
3 Belgium	1962	.86	.86	.71	.81*	-.11*	.34*	-.24*	-.20*	.26*
4 Brazil	2218	.81	.85	.63	.69*	.14*	.23*	-.07	-.08	.25*
5 Czech Republic	2698	.84	.85	.75	.74*	.08*	.51*	-.22*	-.17*	.48*
6 Denmark	2087	.87	.86	.77	.78*	.10*	.65*	-.18*	-.16*	.48*
7 Finland	2576	.84	.93	.80	.71*	.28*	.70*	-.06	.04	.46*
8 Germany	2502	.87	.90	.81	.79*	-.12*	.62*	-.45*	-.41*	.60*
9 Hungary	2550	.86	.87	.67	.76*	.08*	.43*	-.15*	-.17*	.49*
10 Iceland	1720	.86	.91	.78	.75*	.31*	.68*	-.09*	.08	.40*
11 Ireland	2041	.86	.87	.79	.79*	-.11*	.53*	-.22*	-.44*	.47*
12 Italy	2678	.86	.88	.81	.73*	-.06*	.49*	-.21*	-.35*	.62*
13 Korea,	2705	.78	.89	.68	.77*	.13*	.58*	-.19*	-.04	.48*
14 Latvia	1920	.88	.85	.66	.61*	.09*	.37*	-.20*	-.14*	.47*
15 Liechtenstein	153	.83	.85	.76	.76*	-.06	.41*	-.27*	-.37*	.55*
16 Luxembourg	1441	.88	.88	.75	.74*	.01	.40*	-.30*	-.27*	.57*
17 Mexico	2275	.82	.83	.55	.76*	.52*	.14*	-.03	-.10*	.20*
18 Netherlands	1282	.84	.89	.74	.84*	-.07*	.89*	-.75*	-.25*	.34*
19 New Zealand	1809	.88	.89	.80	.80*	-.07*	.80*	-.38*	-.53*	.68*
20 Norway	2050	.88	.90	.74	.73*	.14*	.72*	-.17*	-.20*	.59*
21 Portugal	2378	.89	.87	.73	.79*	.06*	.55*	-.28*	-.18*	.48*
22 Russia	3398	.84	.87	.67	.68*	.29*	.22*	.12*	-.16*	.49*
23 Sweden	2282	.85	.88	.76	.81*	.18*	.58*	-.20*	-.08*	.37*
24 Switzerland	2982	.88	.88	.76	.76*	-.20*	.54*	-.39*	-.28*	.42*
25 United Kingdom	1211	.87	.88	.82	.82*	-.12*	.69*	-.27*	-.38*	.45*
26 United States	1642	.89	.86	.76	.84*	.11*	.42*	-.15*	-.20*	.55*
Mean		**.86**	**.88**	**.74**	**.76**	**.06**	**.51**	**-.22**	**-.21**	**.47**
SD		**.03**	**.02**	**.07**	**.05**	**.17**	**.18**	**.16**	**.14**	**.12**
Median		**.86**	**.88**	**.76**	**.74**	**.08**	**.52**	**-.20**	**-.20**	**.48**
25th %tile		**.84**	**.86**	**.70**	**.76**	**-.07**	**.41**	**-.27**	**-.30**	**.40**
75th %tile		**.88**	**.89**	**.79**	**.79**	**.14**	**.66**	**-.15**	**-.13**	**.55**

Note. MAch =Math Achievement, VAch = Verbal Achievement, MSC = Math Self-concept, VSC = Verbal Self-concept, Uniq = uniqueness. All parameter estimates are present in completely standardized form. Results for each country are based on MG2 (Table 2) in which only factor loadings were constrained to be equal. The "total" results based on all 26 countries are based on Model TG1 (Table 2) in which only uniquenesses were allowed to differ from country to country.

An important feature of the relations between multidimensional achievement scores and multidimensional academic self-concept scores is that academic self-concept scores are substantially less highly correlated – more highly differentiated – than the corresponding academic achievement scores. More specifically, the I/E model predicts that correlations between math and verbal achievement scores are substantial and substantially larger than those between math and verbal self-concept. Although support for the I/E model does not require math and verbal self-concept measures to be uncorrelated, much of the research reviewed earlier has found these two self-concept scores to be nearly uncorrelated. In order to evaluate the cross-cultural generalizability of this pattern of results, the two correlations were presented separately for each country in Table 3 (based on results from Model MG3 in which factor loading and factor variances are invariant across countries, but factor correlations are not). Consistent with a priori predictions, in every country the correlation between the two self-concept scores (M = .06, SD = .17) was consistently much smaller than the correlation between the two achievement scores (M = .76, SD = .07).

Conclusion and Discussion

Results of this study have potentially important theoretical implications for social comparison theory. The theoretical basis for the I/E model extends social comparison theory, positing an internal comparison process in addition to the more typical external comparison process. Specifically, students not only use the performances of other students to form their self-concepts in a particular school subject (the external comparison process), they also use their own performances in other school subjects as a second basis of comparison (the internal comparison process). Although there is clear support for this pattern of results in relation to academic achievements and academic self-concepts, it is also relevant to ask whether similar frame of reference effects exist in other areas as well. We suggest that the implications probably have much broader generality. The critical feature of the internal comparison frame of reference is the use of accomplishments in one arena as a basis of comparison for evaluating accomplishments in another arena.

The extreme domain specificity of academic self-concepts that led to the development of the I/E model also has practical implications for teachers and parents, and for educational practice. Teachers, in order to understand the academic self-concepts of their students in different content areas, must understand the implications of the I/E model. When teachers were asked to infer the self-concepts of their students (see discussion by Marsh & Craven, 1997), their responses reflected primarily the external comparison process so that teacher's inferences were not nearly so domain specific as responses by their students; students who are bright in one area tend to be seen as having good academic self-concepts in all areas whereas students who are not bright in one area are seen as having poor academic self-concept in all areas. Similarly, Dai (2002) reported that inferred self-concept ratings by parents reflected primarily the external comparison process typically emphasized in social comparison research, but not the internal comparison process that is the unique feature of the I/E model. In contrast to inferred self-concept ratings by significant others (teachers and parents), student's academic self-concepts in different domains are extremely differentiated. Hence, understanding the implications of the I/E model will allow significant others to better understand children and to infer childrens' self-concepts more accurately. Thus, for example, our results demonstrate that even bright students may have an average or below average self-concept in their weakest school subject that may seem paradoxical in relation to their good achievement (good relative to other students, but not their own performance in other school subjects). Similarly, even poor students may have an average or above average self-concept in their best school subject that may seem paradoxical in relation to their below-average achievement in that subject. Particularly for poorer students, understanding these principles should assist teachers and parents to give positive feedback that is credible to students.

References

Abu-Hilal, M. M. (April, 2002). Frame Of Reference Model Of Self-Concept And Locus Of Control: A Cross Gender Study In The United Arab Emirates. Paper presented at the 87th annual convention of the American Educational Research Association, New Orleans, LA, April 1-5, 2002

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

Bentler, P. M. (1988). Theory and implementation of EQS. A structural equations program. Los Angles: BMDP Statistical Software.

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

Byrne, B. (1996). Measuring self-concept across the life span: Issues and instrumentation. Washington, DC: American Psychological Association.

Byrne, B. M. (1998). Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications and programming. Mahwah, NJ: Erlbaum.

Byrne, B. M. (in press). Measuring self-concept across culture: issues, caveats, and practice In H. W. Marsh, R. G. Craven & D. McInerney (Eds.), International Advances in Self Research (Vol. 1). Greenwich, CT: Information Age.

Dai, D-Y. (2001). A comparison of gender differences in academic self-concept and motivation between high-ability and average Chinese adolescents. Journal of Secondary Gifted Education, 13 , 22-32.

Dai, D.Y. (2002). Incorporating parent perceptions: A replication and extension study of the internal-external frame of reference model of self-concept development. Journal of Adolescent Research, 17, 617-645.

Joreskög, K.G. and Sorböm., D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. Chicago: Scientific Software International.

Kong, C. (2000). Chinese students' self-concept: Structure, frame of reference, and relation with academic achievement. Dissertation Abstracts International Section A: Humanities & Social Sciences, 61, 880.

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.

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

Marsh, H. W. (1989). Sex differences in the development of verbal and math constructs: The High School and Beyond study. American Educational Research Journal, 26, 191-225.

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

Marsh, H. W. (1990b). Self Description Questionnaire, II. San Antonio, TX: The Psychological Corporation. (Reprinted, 1992, University of Western Sydney)

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. (1994a). Confirmatory factor analysis models of factorial invariance: A multifaceted approach. Structural Equation Modeling, 1, 5-34.

Marsh, H. W. (1994b). Using the National Educational 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., Balla, J. R., & Hau, K. T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical processes. In G. A. Marcoulides & R. E. Schumacker (Ed.), Advanced structural equation modeling techniques (pp. 315-353). Hillsdale, NJ: Erlbaum.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness of fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103, 391-410.

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., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), Handbook of classroom assessment: Learning, achievement, and adjustment (pp. 131-198). Orlando, FL : Academic Press.

Marsh, H. W., & Köller, O. (in press). Unification of two theoretical models of relations between academic self-concept and achievement. In H. W. Marsh, R. G. Craven & D.

Marsh, H. W., Kong, C-K, Hau, K-T (2000). Longitudinal Multilevel Modeling of the Big Fish Little Pond Effect on Academic Self-concept: Counterbalancing Social Comparison and Reflected Glory Effects in Hong Kong High Schools. Journal of Personality and Social Psychology, 78, 337-349.

Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and English constructs. American Educational Research Journal, 35, 705-738.

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

Organisation of Economic Development (OECD, 2001a). Knowledge and Skills for Life: Results from the first OECD programme for international student assessment (PISA) 2000. OECD: Paris, France.

McInerney (Eds.), nternational. Advances in Self Research (Vol. 1). Information Age, Greenwich, CT.

Organisation of Economic Development (OECD, 2001b). PISA international data base (Technical Report. OECD: Paris, France

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

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

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.

Skaalvik, E. M. & Valas, H. (2001). Achievement and self-concept in mathematics and verbal arts: A study of relations. In R J. Riding & S. G. Rayner (Eds), Self perception (pp. 221-238). Westport, CT: Ablex Publishing.

Williams, J. E., & Montgomery, 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 & Psychological Measurement, 59, 431-450.