Ghana Journal of Education: Issues and Practice (*GJE*)

NYANSAPO – "Wisdom Knot"

Symbol of wisdom, ingenuity, intelligence and patience

Ghana Journal of Education: Issues and Practices (GJE)
Vol. 4, December 2018, pp. 44 – 59

Internal/External Frame of Reference: Exploring the Paradoxical relations between Mathematics and Self-Concept across 29 Countries

Emmanuel Adu-tutu Bofah & Forster D. Ntow [*]

College of Education Studies, University of Cape Coast, Ghana

[*]Corresponding author's email address: *fntow@ucc.edu.gh*

Abstract
Students' self-concept is an important construct in explaining achievement-related outcomes. The study examined the theoretical and methodological issues underpinning the Internal/External (I/E) frame of reference model. This model posits a paradoxical relation between "distinct" school subjects, for example, mathematics and verbal. Also, achievement in each domain is deemed to positively affect self-concept in the matching domain but negatively in the nonmatching domain. The investigation is based on 29 countries (N= 181,745) using the TIMSS–2011 data set. The data supported the assumptions associated with the I/E model. Result indicates a negative achievement effect on non-corresponding self-concepts (internal) and positive effects achievement on the corresponding self-concepts (external). The findings contribute to a better understanding of how students form self-concept across domains cross-culturally.

Key words: Internal/external frame of reference effect, mathematics self-concept, mathematics achievement, cross-cultural comparison.

Introduction

Self-concept theories extend across several branches (Wang & Lin, 2008). One such branch concerns the frame of reference effect - the context or standards against which people judge their own accomplishments and failures (Marsh, 2007). More often than not, individuals evaluate their own performance in comparison with the performances of others through social comparison processes (Marsh &

Hau, 2004, p. 57). In most cases, these comparisons are made within individual's immediate context (social comparisons; e.g., classmates in our school or class: external frame of reference (E)), or between domains— one school subject can serve as a frame of reference for another school subject (internal frame of reference (I)) (Marsh, 2007; Marsh et al., 2015; Möller, Helm, Muller-Kalthoff, Nagy, & Marsh, 2015). As Parker, et al. (2014) argue, " for self-concept, students use normative judgments about their ability and social comparison processes with reference to their peers, but also internal comparisons of their performance in one academic domain relative to other academic domains" (Marsh, 2007; Parker et al. 2014, p. 32;). Marsh et al. (2015) emphasized "… perceptions of the self cannot be adequately understood if the role of frames of reference is ignored" (p. 425). Moreover, evidence suggests that self-concept operates differently across cultures (Bofah, 2015; Bofah & Hannula, 2015; Chiu & Klassen, 2010; Marsh & Hau, 2004).

The theoretical rationale for I/E model

The Internal/External (I/E) model was postulated by Marsh (1986) to explain why extremely distinct school subjects, for example, mathematics and verbal (native language) self-concepts are nearly uncorrelated whereas their corresponding areas of achievement correlate substantially. The model posits that academic self-concept in a particular domain (e.g., mathematics or verbal self-concepts) is formed in relation to two comparison processes or frames of reference (Marsh, 2007; Marsh et al., 2015; Marsh & Hau, 2004; Möller et al., 2015); the internal and external frames of reference at the individual level. External comparison is a psychological process in which students compare their own achievement with other students (e.g., school, class) in a particular school domain (e.g., mathematics). The Internal frame of reference is a comparison process in which students compare their own achievements in one particular domain (e.g., mathematics) with that in other domain (e.g., science). Tests of the I/E model are normally examined when mathematics and verbal achievements are regressed on mathematics self-concept (MSC) and verbal self-concept (VSC). The model used to test the external comparison is a horizontal path leading from mathematics achievement (MAch) to MSC and from verbal achievement (VAch) to VSC, and are predicted to be substantial and positive. The model used to examine the internal comparison is cross paths leading from mathematics achievement to verbal self-concept and from verbal achievement to mathematics self-concept and is predicted to be negative (see Figure 1A).

Extant literature have provided evidence supporting the I/E model across cultures (M.-S. Chiu, 2008; Marsh & Hau, 2004). Most studies on the I/E model focus on the two distinctly different domains; mathematics and verbal skills (native language) (e.g., Dickhäuser, 2005; Marsh & Hau, 2004). Additionally, cross-cultural generalisability of the l/E model in domains such as mathematics and verbal skills (Marsh & Hau, 2004) and between science and mathematics (M.-S. Chiu, 2008) have also been examined. Möller, Pohlmann, Koller and Marsh, (2009) also established the I/E through a meta-analysis study for the domains of mathematics and verbal skills across different age groups, gender, and country.

Extension of the I/E model: Mathematics and Science

Marsh and Yeung (2001) argue that limiting the I/E model to one numerical domain (usually mathematics) and verbal (native language) domains is not inherent to the logic of the I/E model. Extending the logic of the original I/E model to other domains will help construct stronger measures of these internal and external comparison processes (Marsh & Yeung, 2001; Möller, Streblow, Pohlmann, & Köller, 2006). The present study takes up this challenge using the mathematics and science constructs of the TIMSS 2011 cross-national study for 29 countries. The purpose of this study is to explore whether the I/E model can be extended to science and mathematics, two school subjects perceived to be highly related. The present study, therefore, extends previous research on the I/E models based almost completely on two distinct school domains (verbal and mathematics) to mathematics and science. Such an approach has the potential to help in the generalization of the I/E model. Moreover, this study will help clarify the suggestion that mathematics and science are either two distinct domains that are complementary or supplementary to each other in terms of national curricula, knowledge types, and student perceptions (M.-S. Chiu, 2008, 2012). The research questions are:

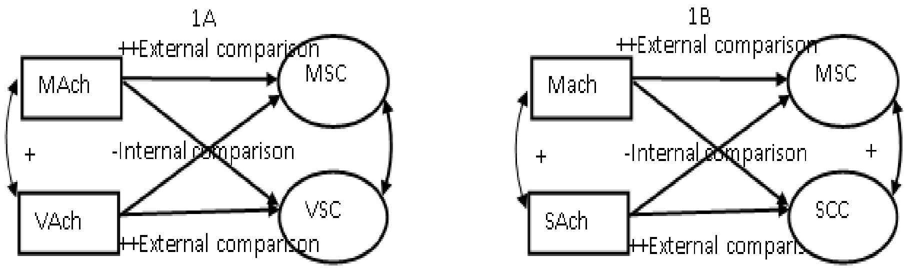
1. is the internal/external (I/E) frame of reference model supported when extended to other school subjects such as mathematics and science?

2. how does the I/E model perform cross-culturally?

Methodology

The present study hypothesized that there will be support for the I/E model across the general sample. More specifically, each of mathematics and science achievements will have a positive effect on the matching self-concept domain (external frame of reference: the horizontal paths in Figure 1B), but negative effects on nonmatching domains—mathematics achievement will have a negative effect on science self-concept and science achievement having a negative effect on mathematics self-concept (internal frame of reference: the cross-paths in Figure 1B). Moreover, mathematics and science achievement is expected to be positively correlated as well as mathematics and science self-concepts. The expected pattern of results is depicted in Figure 1B.

Figure 1. The original internal/external frame of reference (I/E) model relating verbal and math achievements to verbal and math self-concepts (1A).



In the present investigation, we evaluated the generalizability of this model to science domain, relating science and math achievements to science and math self-concepts (1B). Coefficients indicated to be "++", "-" are predicted to be positively high and negatively low. To avoid cluttering, only paths are shown.

Data Source

Data were obtained from students who participated in the Trends in International Mathematics and Science Study (TIMSS) 2011 study. For the present study, participants were eighth-graders from 29 countries (N = 181,745 see Table 2). See Martin and Mullis (2012) for detailed TIMSS sampling and method procedures.

TIMSS - 2011 students' self-concept in mathematics (MSC) and science (SCC) scales were created based on students' degree of agreement to each of five statements. Each scale was measured on a 4-point Likert response format: Agree a lot (1), Agree a little (2), Disagree a little (3), Disagree a lot (4). In the present study, the scores on two of the items were reverse-coded so that a higher value corresponds to a higher response. The items on the MSC/SCC are the following: 1) I usually do well in mathematics/Science, 2) I learn things quickly in mathematics/Science, 3) I am good at working out difficult mathematics/Science problems, 4) Mathematics/Science is more difficult for me than for many of my classmates [reverse coded], and 5) Mathematics/Science is not one of my strengths [reverse coded].

TIMSS - 2011 reported students' achievement in terms of five plausible values—random numbers drawn from the distribution of scores that could be reasonably assigned to each individual (Martin & Mullis, 2012). All reported five plausible achievement measures in mathematics and science were used in this study (See Martin and Mullis (2012) for discussion on the use of plausible values).

Data Analysis - Model evaluation and estimation criteria

Data analyses were conducted by means of structural equation modelling (SEM) with Mplus 7.4. SEM analyses were applied to conduct confirmatory factor analyses (CFAs), path analyses with plausible and latent variables and multiple-groups analyses. Missing data were addressed with the Mplus feature of multiple imputations with five imputed data sets (e.g., Asparouhov & Muthén, 2010). To ascertain the model fit, emphasis was placed on the comparative fit index (CFI; normed along a 0-to-1 continuum with values over .90 representing an adequate fit), the root-mean-square error of approximation (RMSEA; values less than .08 are indicative of a reasonable fit) (West, Taylor, & Wu, 2012) and the chi-square test statistic (χ^2: for informative purposes only because of its sensitivity to large sample size). These cut-off standards may be specific to particular models (complex models with large sample size) and data sets and using fit indices for interpreting acceptable model fits are only rough guidelines (West, Taylor, & Wu, 2012). The complex design associated with TIMSS data was accounted for by incorporating the clustering variable, and students' sampling weights (weighting variable supplied with the data) in the analysis.

Moreover, following work on method effects associated with parallel and negatively worded items by Marsh et al. (2013) (see also Bofah & Hannula, 2015; M.-S. Chiu, 2008; Marsh et al., 2014), correlated uniqueness accounting for method effects relating to parallel and negative item wording was incorporated into all the models.

Cultural equivalence test of the I/E model

Establishing measurement invariance is regarded as an important condition for any construct validity or theoretical generalisation in cross-cultural research. In fact, most research on cross-cultural comparison advocate the use of measurement invariance to ease cross-cultural generalisability of the measured models (e.g., Vandenberg & Lance, 2000). There are several levels of measurement invariance. Three levels are of significance in cross-cultural comparison: configural (all the parameters are freely estimated across the groups), metric (the factor loadings are equally held across the group), and scalar (invariance constraints are placed on the measurement intercepts and the factor loadings) invariance. However, if mean comparisons are not the objective, as in the present study, then configural and metric are of merit.

Multigroup CFAs were used to examine if the measurement and structural models (I/E model) are invariant across the 29 educational systems/cultures. For the measurement model the configural (MG1) and metric (MG2) invariance were examined across the groups. To ascertain whether the I/E is invariant across the 29 educational systems, models whereby all estimates of the I/E model were constrained to be equal across the educational systems were compared to a model with all constraints freely estimated. Due to restrictions of space, detailed analyses are not presented in this paper.

Results

Using the overall sample (TG in Table 1 and overall data in Table 2), the I/E model was supported. The path estimates were consistent with the I/E framework (Figure 1B) that is, negative path from mathematics achievement to science self-concept ($\beta = -.564$) and from science achievement to mathematics self-concept ($\beta = -.197$), a positive path from mathematics achievement and science achievement to mathematics self-concept (β = .310) and science self-concept (β = .539) respectively. There was a positive correlation between mathematics and science achievement (r = .880) and between science and mathematics self-concept (r = .426).

Multigroup CFA was used to examine if the I/E model is invariant across the 29 countries. The results of the country equivalence tests of the I/E model were acceptable. The RMSEA indicated a reasonable fit, whereas CFI and TLI indicated a slightly below reasonable fit model (see Table 1). The results of the country invariance test indicated a less adequate fit with increasing constraints on parameter estimates. For cultural invariance of the I/E model, comparison was made between two nested models (see Table 1), MG3 (freely estimating all parameter estimates) and MG6 (constraining all estimated parameter estimates to be equal across group). There was support for the invariance of the I/E model across countries because of the small change in the fit indices of the compared models (ΔCFI = .014, ΔTLI = .002, ΔRMSEA = .000; though $\Delta\chi^2_{188}$ = 5867.536, $p < .05$). Other model comparisons, MG4 vs. MG5 and MG4 vs. MG6 also indicated a similar outcome. Moreover, there were variations in parameter estimates with respect to each country. A thorough look at each country's parameters indicated that a fit to the I/E model occurred in 24 out of the 29 countries. Five countries deviated from the I/E model (Table 2). Four partially supported and one contrasted the I/E model. Since there were no significant changes in model indices when constraints were imposed on the parameter estimates for the cultural equivalent test, separate analysis was done for each country in verifying the I/E model (Table 2).

Confirmatory factor analysis (CFA) with a two-factor model (M1: Table 1), with each domain representing a unique dimension did not fit the data. However, controlling for method effects associated with parallel and negatively worded items (M3) indicated a substantial improvement in goodness-of-fit. Moreover, there was support for configural (MG1) - reasonable to think that the factor structure is applicable across all 29 countries, and metric invariance (MG2) - one could conclude that the constructs are manifested in the same way in each of the groups - across the 29 country. All analyses subsequently reported are based on metric invariance model. All estimates of the factor loadings are constrained to be equal across the 29 countries.

Models MG1-MG6 examined whether the I/E model fit the 29 countries by setting certain combinations of invariants and freely estimating some parameters across the countries (See Table 1).

Composite reliabilities estimates for the MSC and SCC scales reached the acceptable value of .80, but in few cases fell below the acceptable value of .60. These are average results over five data sets due to the use of plausible values in the computations (See Table 2). Reliabilities were generally lower for the SCC construct than for the MSC. The lower reliabilities may attenuate the validity of the interpretations of the results and weaken the statistical power as well as the effect sizes (Schmidt & Hunter 1996). This necessitated the use of latent-variable models that accounted for unreliability, bias, and measurement errors (Cole & Preacher, 2014).

Table 1: Model TG examines whether the I/E model fit the total group

Model	χ^2	*df*	CFI	TLI	RMSEA	Model description
TG	19168.672	41	.936	.899	.051	Total group
M1	68909.136	34	.770	.695	.106	Two factor model
M2	13813.936	28	.954	.926	.052	includes CUs for negative item effect
M3	4268.378	25	.986	.974	.031	CUs for both negative and matching item effect
Multigroup CFA						
MG1	6264.252	725	.984	.971	.035	Inv =none
MG2	15487.215	949	.957	.941	.050	Inv =FL
Multigroup I/E model						
MG3	51351.300	1637	.872	.853	.070	Inv =none
MG4	51351.298	1637	.872	.853	.070	Inv = FL
MG5	56900.881	1749	.858	.847	.071	Inv = FL, PC
MG6	57218.836	1805	.858	.851	.070	Inv =FL, PC, FC, AC

Note. TG = total group; MG = multiple group (or multigroup); CFA = confirmatory factor analysis; Inv = invariant; FL = factor loadings; FC = factor covariance; AC = correlation between mathematics and science achievement; PC = path coefficients, CU=Corrected uniqueness.

Table 2: Composite reliabilities estimates for the MSC and SCC scales for 29 Countries

		Reliabilities		Path coefficient				Correlation	
Country	N	MSC ω	SCC ω	MAch to MSC	MAch to SCC	SAch to MSC	SAch to SSC	MSC - SSC	MAch - SAch
Australia	7556	.858	.836	.752	-.271	-.200	.661	.342	.851
Bahrain	4640	.654	.548	.636	-.302	-.260	.570	.371	.878
Botswana	5400	.648	.614	.989	-.226	-.953	.360	.330	.874
Chile	5835	.773	.711	.744	-.589	-.336	.699	.087	.846
Chinese Taipei	5042	.880	.850	.761	-.281	-.175	.708	.457	.877
Palestinian National Authority	7812	.601	.595	.649	-.**<u>069</u>**	-.295	.407	.450	.891
Ghana	7323	.515	.457	.397	-.118	-.254	.326	.369	.804
Honduras, Republic of	4418	.616	.536	.370	-.**<u>056</u>**	-.214	.**<u>116</u>**	.362	.824
Hong Kong, SAR	4015	.805	.751	.614	-.388	-.229	.577	.223	.831
Iran, Islamic Republic of	6029	.714	.671	.646	-.**<u>061</u>**	-.267	.329	.413	.859
Israel	4699	.773	.792	.659	-.345	-.239	.665	.116	.861
Italy	3979	.876	.804	.692	-.283	-.165	.542	.293	.827
Japan	4414	.824	.806	.691	-.098	-.123	.569	.478	.838
Jordan	7694	.543	.459	.788	-.161	-.457	.452	.487	.892
Korea, Republic of	5166	.861	.859	.814	-.**<u>044</u>**	-.150	.622	.524	.841
Malaysia	5733	.600	.599	.385	-.445	-.170	.589	.482	.803
Oman	9542	.470	.462	.665	-.**<u>043</u>**	-.326	.398	.465	.888

Table 2 Cont'd: Composite reliabilities estimates for the MSC and SCC scales for 29 Countries

		Reliabilities		Path coefficient				Correlation	
Country	N	MSC ω	SCC ω	MAch to MSC	MAch to SCC	SAch to MSC	SAch to SSC	MSC - SSC	MAch - SAch
New Zealand	5336	.821	.790	.797	-.379	-.331	.651	.287	.839
Norway	3862	.871	.814	.841	-.119	-.230	.488	.430	.821
Qatar	4422	.622	.614	.481	-.304	-.217	.590	.323	.873
Saudi Arabia	4344	.652	.587	.602	-.214	-.151	.520	.398	.845
Singapore	5927	.845	.837	.918	-.491	-.557	.719	.166	.877
South Africa	11969	.570	.533	.540	-.194	-.546	.199	.349	.856
Thailand	6124	.604	.574	.415	-.182	-.384	.182	.512	.827
United Arab Emirates	14089	.659	.622	.671	-.298	-.362	.518	.274	.869
Tunisia	5128	.626	.550	.621	-.201	-.276	.451	.141	.824
Turkey	6928	.779	.709	.850	-.166	-.364	.551	.412	.896
United States	10477	.824	.791	.738	-.225	-.319	.556	.134	.835
England	3842	.815	.823	.848	-.145	-.438	.433	.342	.846
Overall data	181745	.775	.757	.310	-.564	-.197	.539	.426	.880

Note: MAch = math achievement, SAch = science achievement, SCC = science self-concept, MSC = mathematics self-concept, ω = composite reliability. The estimates underlined are not significant at the .05 leve

Discussion

The purpose of this study was to examine the I/E frame of reference model on students' self-concept in mathematics and science. The use of science and mathematics in exploring the I/E has helped broaden the generalisability of the I/E model across other school domains. The findings of this study clearly support as well as challenge the foundations of the I/E theories, and provide new ways of looking at the theory. As such, the present study results are consistent with findings of Marsh and Hau's (2004) PISA study, Möller et al.'s (2009) meta-analysis, and Marsh et al.'s (2014) provide the strongest support for the generalisability of the I/E frame of reference effect model. That is, the internal/external frame of reference (I/E) model posits a paradoxical relation between "distinct" school subjects, for example, mathematics and verbal (native language). Also, achievement in each domain positively affects self-concept in the matching domain but negatively in the nonmatching domain. Furthermore, the choice of mathematics and science as the subjects of choice in this paper helped test the generalisability of the I/E model beyond mathematics and verbal skills used to be the norm in a number of the papers that were reviewed.

Additionally, the measurement nature of self-concepts model and the I/E model was supported by the overall data. A series of confirmatory factor analyses (CFAs) revealed the need to control for the method effects associated with such items. Although, composite reliability estimates were very low for some countries, these reliability estimates emphasize the need for latent variable models, such as those used in the present investigation (Marsh et al., 2014).

In this study, 24 out of the 29 countries supported the I/E model. This indicates that students in the 24 countries have a clear distinction between their mathematics and science self-concepts, and their mathematics and science abilities, although there was low to medium correlation between their mathematics and science self-concept and a high correlation between their mathematics and science achievement. However, five countries did not fully support the I/E model. Out of the five countries, four partially supported the I/E model and one contrasted it. For these four countries, the internal frame of reference path from mathematics achievement to science self-concept was not statistically significant. In the case of the fifth country, the internal frame of reference estimate between mathematics achievement and science self-concept as well as the external frame of reference estimate between science achievement and self-concept were not statically significant. Interestingly, three out of the five countries partially supporting the I/E model were countries in the Middle East. The five countries that did not conform to the I/E model may give an indication of the influence of local school systems or culture on self-concept and achievement. With the 24 countries supporting the I/E model, students' self-perception as to whether mathematics and science are supplementary or different domain are clear. In these countries, mathematics and science are two distinct domains. The other five countries reported partial support for mathematics and science being two distinctive domains. Similar outcomes were reported in M.-S. Chiu (2008) and Marsh et al. (2014) in their study. The fact that countries differ to some extent on the I/E model suggests that research and theorizing that integrate cross-cultural perspectives are crucial to the establishment of more useful and universal theories (e.g., van de Vijver & Leung, 2000). This finding gives some indication of possible variations in how students from different cultures form self-concepts and add to the literature on cross-cultural studies and the validity of studies carried across multiple nations.

One important finding of this study is that social comparison processes influence internal and external self-concepts. Students with good grades in mathematics are less likely to think that they are good in science and vice versa. Students are more likely to perceive themselves to be more talented in mathematics than science when they perceive their mathematics achievement to be better than those of their classmates (Dickhäuser, 2005).

A limitation of the study is that the model could not account for the reciprocal determinism between mathematics and self-concept. This is because cross-sectional data used in modelling a reciprocal analysis is very problematic (Bofah, 2015). Moreover, the indices for the final model and some of the reliability estimates of the self-concept constructs were below acceptable scores. Although, method effect associated with negatively phrased and parallel items were controlled, the factor loadings of the negatively phrased items were very low and varied across all models tested. This supports the notion that responses to negatively phrased items are culture-specific (see Bofah & Hannula, 2015). Notwithstanding this challenge, a strength of the present study is that the TIMSS data used here are nationally representative samples of students carefully constructed, and was consistently measured for a diverse set of countries (Marsh et al., 2014). Thus, the conclusion is that support for the I/E model can be generalised cross-culturally through the use of a more robust approach and stronger data set as reported in this paper. Another advantage in the approach adopted in this study as compared to other similar studies (e.g., Chiu, 2012) is that we included both latent variables that controlled for method factors and measurement error and all five achievement plausible values as discussed in TIMSS documentation.

References

Asparouhov, T., & Muthén, B. O. (2010). *Multiple imputation with Mplus: Mplus technical appendices.* Los Angeles, CA: Muthen & Muthen. Retrieved from http://statmodel2.com/download/Imputations7.pdf.

Bofah, E. A. (2015). Reciprocal determinism between students' mathematics self-concept and achievement in an African context. In K. Krainer & N. Vondrová (Eds.), *CERME 9-Ninth Congress of the European Society for Research in Mathematics Education* (pp. 1688–1694). Prague, Czech Republic. Retrieved from https://hal.archives-ouvertes.fr/hal-01287995.

Bofah, E. A., & Hannula, M. S. (2015). TIMSS data in an African comparative perspective: investigating the factors influencing achievement in mathematics and their psychometric properties. *Large-Scale Assessments in Education, 3*(1), 4.

Chiu, M.-S. (2008). Achievements and self-concepts in a comparison of math and science: exploring the internal/external frame of reference model across 28 countries. *Educational Research and Evaluation*, *14*(3), 235–254.

Chiu, M.-S. (2012). The internal/external frame of reference model, big-fish-little-pond effect, and combined model for mathematics and science. *Journal of Educational Psychology, 104*(1), 87–107.

Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction*, *20*(1), 2–17.

Cole, D. a, & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods, 19*(2), 300–15.

Dickhäuser, O. (2005). A fresh look: Testing the internal/external frame of reference model with frame-specific academic self-concepts. *Educational Research, 47*(3), 279–290.

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, *23*(1), 129–149.

Marsh, H. W. (2007). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. Leicester, UK: British Psychological Society.

Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J. S., Abdelfattah, F., Leung, K. C., … Parker, P. (2013). Factorial, convergent, and discriminant validity of TIMSS math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. *Journal of Educational Psychology, 105*(1), 108–128.

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J. S., Abdelfattah, F., Nagengast, B., … Abu-Hilal, M. M. (2014). The internal/external frame of reference model of self-concept and achievement relations: Age-Cohort and cross-cultural differences. *American Educational Research Journal*, *52*(1), 168–202.

Marsh, H. W., & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, *96*(1), 56–67.

Marsh, H. W., Möller, J., Parker, P., Xu, M. K., Nagengast, B., & Pekrun, R. (2015). Internal/External frame of reference model. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed., Vol. 12, pp. 425–432). Amsterdam, Netherlands: Elsevier.

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*(3), 389–420.

Martin, M. O., & Mullis, I. V. S. (Eds.). (2012). *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from http://pirls.bc.edu/methods/index.html

Möller, J., Helm, F., Muller-Kalthoff, H., Nagy, N., & Marsh, H. W. (2015). Dimensional comparisons and their consequences for self-concept, motivation, and emotion. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed., Vol. 6, p. ¨430-436). Amsterdam, Netherlands: Elsevier.

Möller, J., Pohlmann, B., Koller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research, 79*(3), 1129–1167.

Moller, J., Streblow, L., Pohlmann, B., & Köller, O. (2006). An extension to the internal/external frame of reference model to two verbal and numerical domains. *European Journal of Psychology of Education, XXI* (4), 467–487.

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology*, *34*(1), 29–48.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1*(2), 199–223.

van de Vijver, F. J. R., & Leung, K. L. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology, 31*(1), 33–51.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70.

Wang, J., & Lin, E. (2008). An alternative interpretation of the relationship between self-concept and mathematics achievement: Comparison of Chinese and US Students as a context. *Evaluation & Research in Education*, *21*(3), 154–174.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). New York, NY: The Guilford Press.