

Chapter III

METHOD

Overview

This study used data from the TIMSS-R 1999. The methodology used in the TIMSS-R (1999) study is presented in this section. TIMSS participants and procedures are described. Measures were developed of autonomy support, intrinsic motivation, external regulation, introjected regulation, and math self-concept. Data analysis for the present study is described in the last section.

TIMSS Participants & Procedures

Participants in the TIMSS-R (1999) study were 9,072 eighth-grade students, ranging in age from 13 to 14 years, from 150 sampled schools in the U.S.. The subjects had an average age of 14.2 years. 50.2% were women, 49.8% were men, and 90.4% were from English-speaking backgrounds. The education levels of their mother and father ranged from completion of primary school to a bachelor's degree.

The TIMSS-R (1999) used a two-stage sampling procedure to ensure a representative sample of U. S. students (Mullis et al, 2000). After the researchers

randomly selected a sample of schools, they randomly selected one or two math classrooms in the sampled schools - each student was a member of one of those math classes. A sample of 150 schools resulted in 95% confidence limits for school-level and classroom-level mean estimates precise to within $\pm 16\%$ of their standard deviations (Mullis et al., 2000). Sampling methods are described in more detail in *TIMSS 1999 School Sampling Manual*, the *Survey Operations Manual*, and *School Coordinator Manual* (Mullis et al., 2000).

Subjects were given a TIMSS 1999 student background questionnaire in the sampled classes. They were asked to respond to the student questionnaire measuring students' home backgrounds, how they spend their time out of school, their math self-concept, and their attitudes toward math (Mullis et al., 2000).

Teachers were also given a teacher background questionnaire in the sampled classes. They were asked to respond to the teacher background questionnaire measuring teachers' background and training, their instructional practices, the materials used in instruction, the activities students do in class, the use of calculators and computers in math lessons, the role of homework, and the reliance on different types of assessment approaches.

The NRCs and International Study Center staff developed the TIMSS math test, and reviewed the items and scoring guides. They also reviewed item statistics computed to determine the difficulty of each item, the reliability of the scoring of response items, how well items discriminated between high-and low-performing students, and whether there were any biases (Mullis et al., 2000).

Items consisted of five content areas: fractions and number sense; measurement; data representation analysis and probability; geometry; and

algebra. Five performance expectations were covered in the math test: knowing, using routine procedures, investigating and problem solving, mathematical reasoning, and communicating (Mullis et al., 2000).

One-fourth of the questions in free-response format asked students to write short answers. The response items were scored by using two-digit codes with rubrics specific to each item (Mullis et al., 2000). The first digit was designed to represent the correctness level of the response, and the second digit was designed to identify types of strategies (Mullis et al., 2000).

Two independent scorers who did not know the scores assigned by the others scored free response items in one-quarter of the test booklets (Mullis et al., 2000). The degree of agreement between the two scorers was used to measure the reliability of the scoring process. The inter-rater reliability between the two independent scorers was .99, and the inter-rater reliability between the two scorers who agreed on the two-digit code for each item was .98 (Mullis et al., 2000). Item difficulty was more than 95% in the sample as a whole, and item difficulty was less than 25% for multiple-choice items in the sample as a whole (Mullis et al., 2000). The Rasch goodness-of-fit index was less than 0.88 or greater than 1.12 (Mullis et al., 2000).

The Present Study

This section discusses the selection of a modeling subsample for the present study, and describes key variables. In order to use the TIMSS background questionnaire data to test the hypotheses of the present study,

measures were developed of autonomy support, intrinsic motivation, external regulation, introjected regulation, and math self-concept.

In the TIMSS study, there were eight different tests, with some questions occurring in multiple tests. Test items were grouped into 26 mutually exclusive test item “clusters”. There were eight test booklets, each containing seven clusters. Math was emphasized in four booklets, with 34, 37, 37, and 38 minutes for math. Science was emphasized in four booklets, with 18, 19, 19, and 24 for math.

For the present study, we used only four booklets in which math was emphasized to enhance comparability across booklets. We measured math performance in the TIMSS-R 1999 mathematics test with indicators such as first math overall plausible value scores or multiple imputation methods. Each student was administered only a fraction of the mathematics items. A plausible value is an estimate of how each student might have performed if they had been administered the entire set of items. Five plausible values were computed for each student, based on responses to the item set administered and using alternative models based on responses by students with similar characteristics and other items. Examining classroom performance and groups of students provide performance trends. Plausible value score is a measure of total score, but adjusted to make different booklets comparable. The methodology for computing plausible values is described in item response theory.

Autonomy Support

Autonomy support in the classroom was measured using five items from the TIMSS 1999 teacher questionnaire.

In the TIMSS study, participants responded to the stem “In math lessons, how often do students_____?” The stem was followed by five items as follows: “work individually without assistance from the teacher”, “work individually with assistance from the teacher,” “work together as a class with students responding to one another,” “work in pairs or small groups without assistance from the teacher,” and “work in pairs or small groups with assistance from the teacher” (see Table 1). Responses were indicated on a 4-point Likert scale, which ranged from 1(*never/almost never*) to 4 (*every lesson*) (Mullis et al., 2000).

Motivation

Measures were developed of intrinsic motivation, external regulation, and introjected regulation.

Intrinsic motivation was measured using three items from an index of positive attitudes towards mathematics (PATM) of the TIMSS 1999 student questionnaire. Three items from PATM were as follows: “I like mathematics,” “I enjoy learning mathematics,” and “Mathematics is boring¹.”

¹ The response categories for this statement are reversed in constructing the scale.

External regulation was measured using three items from the TIMSS 1999 student questionnaire: “I need to do well in mathematics to get the job I want,” “I would like a job that involved mathematics,” and “I need to do well in mathematics to get into the <secondary school> I prefer.”

Introjected regulation was measured using five items from the student questionnaire: “I need to do well in mathematics to please myself,” “I need to do well in mathematics to please my parents,” “My mother thinks that it is important for me to do well in mathematics at school,” “Most of my friends think it is important for me to do well in mathematics at school,” and “I think it is important to do well in mathematics at school” (see Table 1).

In the TIMSS study, students were asked to rate their motivation towards mathematics on a four-point Likert scale, which ranged from 1 (*strongly agree*) to 4 (*strongly disagree*) (Mullis et al., 2000).

Math Self-Concept

Math self-concept (assessment of personal characteristics regarding math ability) was measured using five items from an index of students' self-concept in mathematics (SCM) of the TIMSS 1999 student questionnaire.

Five items from SCM were as follows: “I would like mathematics much more if it were not so difficult,” “Although I do my best, mathematics is more difficult for me than for many of my classmates,” “Nobody can be good in every subject, and I am just not talented in mathematics,” “Sometimes when I do not understand a new topic in mathematics initially, I know that I will never really understand it,” and “Mathematics is not one of my strengths”(see Table 1).

In the TIMSS study, students were asked to rate their math self-concept on a four-point Likert scale, which ranged from 1 (*strongly agree*) to 4 (*strongly disagree*) (Mullis et al., 2000).

Table 1

List of Items and Scale

Item	Scoring
Autonomy support	
BTBMLES1	"In math lessons, how often do students work individually without assistance from the teacher?" 1(Never) – 4 (every lesson)
BTBMLES2	"In math lessons, how often do students work individually with assistance from the teacher?" 1(Never) – 4 (every lesson)
BTBMLES4	"In math lessons, how often do students work together as a class with students responding to one another?" 1(Never) – 4 (every lesson)
BTBMLES5	"In math lessons, how often do students work in pairs or small groups without assistance from the teacher?" 1(Never) – 4 (every lesson)
BTBMLES6	"In math lessons, how often do students work in pairs or small groups with assistance from the teacher?" 1(Never) – 4 (every lesson)
Intrinsic motivation	
BSBMENJY	"I enjoy learning mathematics." 1(strongly agree) – 4 (strongly disagree)
BSBMLIKM	"I like mathematics." 1 (Like a lot) – 4 (Dislike a lot)
BSBMBORE (-)	"Math is boring." 1(strongly agree) – 4 (strongly disagree)
External regulation	
BSBMJOB	"I need to do well in mathematics to get the job I want." 1 (strongly agree) - 4 (strongly disagree)
BSBMWORK	"I would like a job that involved using mathematics." 1 (strongly agree) - 4 (strongly disagree)
BSBMSCHL	"I need to do well in mathematics to get into the <secondary school> I prefer." 1(strongly agree) – 4(strongly disagree)
Introjected regulation	
BSBMSELF	"I need to do well in mathematics to please myself." 1 (strongly agree) - 4 (strongly disagree)
BSBMPRNT	"I need to do well in mathematics to please my parents." 1 (strongly agree) - 4 (strongly disagree)
BSBMMIP2	"My mother thinks that it is important for me to do well in mathematics at school." 1(strongly agree) - 4 (strongly disagree)
BSBMFIP2	"Most of my friends think it is important for me to do well in mathematics at school." 1(strongly agree) – 4 (strongly disagree)
BSBMSIP2	"I think it is important to do well in mathematics at school." 1(strongly agree) - 4(strongly disagree)
Math self-concept	
BSBMMYT1	"I would like mathematics much more if it were not so difficult." 1(strongly agree) – 4(strongly disagree)
BSBMMYT2	"Although I do my best, mathematics is more difficult for me than for classmates." 1(strongly agree) – 4 (strongly disagree)
BSBMMYT3	"Nobody can be good in every subject, and I am just not talented in mathematics." 1(strongly agree) – 4 (strongly disagree)
BSBMMYT4	"Sometimes when I do not understand a new topic in math initially, I know I will never understand it." 1(strongly agree) – 4 (strongly disagree)
BSBMMYT5	"Mathematics is not one of my strengths." 1 (strongly agree) - 4 (strongly disagree)
Achievement	
BSMMAT01	1 st plausible value

Table 2

Descriptive Statistics for Each Measure

Item	M	SD	Range	Skewness	Kurtosis
Autonomy support	10.14	1.16	5-20	1.087	3.207
BTBMLES1	2.09	0.84	1-4	.59	-.08
BTBMLES2	2.72	0.79	1-4	.05	-.67
BTBMLES4	2.33	0.74	1-4	.63	.18
BTBMLES5	1.99	0.64	1-4	.72	1.69
BTBMLES6	2.28	0.64	1-4	.68	.68
Intrinsic motivation	4.33	1.66	2-8	.51	-.39
BSBMENJY	2.15	0.87	1-4	.44	-.44
BSBMLIKM	2.17	0.92	1-4	.52	.04
BSBMBORE	1.79	0.95	1-4	-.19	-.89
External regulation	5.89	1.86	3-12	.41	-.11
BSBMSCHL	1.48	0.65	1-4	1.30	1.76
BSBMJOB	1.78	0.81	1-4	.78	-.07
BSBKWORK	2.61	0.96	1-4	-.05	-.97
Introjected regulation	8.38	2.13	5-20	.88	1.45
BSBMSELF	1.73	0.81	1-4	.97	.48
BSBMPRNT	1.89	0.80	1-4	.64	-0.03
BSBMMIP2	1.28	0.52	1-4	1.97	4.83
BSBMSIP2	1.38	0.57	1-4	1.45	2.68
BSBMFIP2	2.08	0.72	1-4	.57	.56
Math self-concept	13.49	3.79	5-20	-.21	-.65
BSBMMYT1	2.28	0.96	1-4	.15	-.98
BSBMMYT2	2.71	0.96	1-4	-.36	-.82
BSBMMYT3	2.80	0.98	1-4	-.44	-.80
BSBMMYT4	3.02	0.89	1-4	-.72	-.14
BSBMMYT5	2.66	1.05	1-4	-.23	-1.15
Achievement					
BSMMAT01	495.74	85.90	216.87-813.38	-.06	-0.02

Note. See Table 1 for definitions of the items.

Data Analysis

First, hierarchical linear modeling (HLM) techniques were used to examine individual- and classroom-level effects within a hierarchical structure. Second, structural equation modeling (SEM) analysis with autonomy support treated as student-level was performed to examine 1) if motivational resources predict math performance; 2) if math self-concept explains additional variance in math performance, controlling for the mediating effect of intrinsic motivation on math performance; and 3) if autonomy support predicts math performance through the mediator of math self-concept. Third, multilevel path modeling analysis was performed.

Hierarchical Linear Modeling

Multilevel analysis is used because students are nested within classes, which are nested within schools. In the present study we test variables at only two levels: student and classroom, using HLM analyses.

The general purposes of specific types of HLM Models are as follows: In the one-way ANOVA model with classroom as a random effect, we want to test if there are differences in math achievement among classes. In the means-as-outcomes regression model, we want to test if autonomy support in the classroom significantly affects mean math achievement. In the random-intercept model, we want to test if intrinsic motivation, external regulation, and math self-concept significantly predict

math achievement, while controlling for classroom differences. In the random-coefficients regression model, we want to test if the regression slopes of math performance on these predictors vary from class to class. In the intercept-and slopes-as-outcomes model, we want to test if autonomy support in the classroom accounts for some of the variability in the student-level slopes.

Structural Equation Modeling

Structural equation modeling (SEM) analysis was performed, using the statistical program Mplus (Muthen & Muthen, 1998). SEM is a set of statistical techniques that include confirmatory factor analysis (CFA) and path modeling (Ulman, 2001).

The full model (Figure 2) features four exogenous variables (autonomy support, gender, mother's education, and external regulation) and three endogenous variables (math self-concept, intrinsic motivation, and math performance). Demographic control variables (gender & mother's education) are added as covariates.

In adding demographic control variables, we drew upon the theoretical bases of TIMSS-R (1999) study. Mullis et al. (2000) found that there was a significant gender difference in math self-concept, but that gender was not related to math performance. Mullis et al. (2000) also found that there was a significant gender difference in attitudes towards mathematics. They also found that parents' education was positively related to students' math achievement.

We conducted a series of SEM analyses. Several indices were used to assess model fit: the chi-square statistic, the *comparative fit index* (CFI), the *non-*

normed fit index (NNFI), the *root-mean-square error of approximation* (RMSEA), the *standardized root mean-square residual* (SRMR) (Hu & Bentler, 1999), the Akaike information criterion (AIC), and the Bayesian information criterion (BIC).

The overall fit of the model to the data was examined using the chi-square test. A nonsignificant chi-square indicates the model to be an acceptable fit to the sample data (Bollen & Long, 1993). However, because the chi-square statistic is very sensitive to sample size (Kline, 1998), we examined a χ^2/df ratio to check fit. A ratio of 3 or less indicates a good fit² (Kline, 1998).

The CFI, the NNFI, and the RMSEA were used as incremental fit indices. The CFI compares the lack of fit of the hypothesized model with the independence model (Bentler, 1990). The CFI ranges from 0 (poor fit) to 1 (perfect fit) - cutoff values greater than .95 show acceptable fits to the data (Hu & Bentler, 1999; Schumacker & Lomax, 1996). The NNFI includes a correction for model complexity (Kline, 1998). The NNFI ranges from 0 (poor fit) to 1 (perfect fit) - cutoff values greater than .95 show acceptable fits to the data (Hu & Bentler, 1999). The RMSEA measures the amount of unfitted residuals between the implied and observed covariance matrices (Browne & Cudeck, 1993; as cited in Standage, Duda, & Ntoumanis, 2003). RMSEAs less than .05 represent a "close fit," and values up to .08 show errors of approximation (Browne & Cudeck, 1993). The CFI does not include a penalty for a lack of parsimony, whereas the NNFI and RMSEA include a penalty for a lack of parsimony (Hu & Bentler, 1999).

The SRMR was used as an absolute fit index. The SRMR measures the degree to which the a priori structure fails to reproduce the data - the SRMR

² This recommendation applies to large sample analyses (Kline, 1998).

value should be .08 or less (Hu & Bentler, 1999). When the fit of the model is perfect, the SRMR equals zero (Kline, 1998).

The AIC and the BIC were used to find parsimonious models that best fit the observed data. Even though there are no criterion values on the basis of which it can be inferred if a model fits the data, we can choose the model that results in the lowest value.

We tested a series of nested models to compare one model against another. The hypothesized model can be nested within an alternative model if the alternative model includes a path that is not included in the hypothesized model (Bentler, 1990). We can test the difference in chi-square between the models.

Multilevel Path Modeling

The intraclass correlation value of .53 shows that a substantial proportion of the variance, around 53%, is between classes. Thus, as a first step, we tested the within-level model separately. Muthen (1994) recommends that we run the within- and between-level models separately to determine possible sources of misspecification, since there are no modification indexes in *Mplus* for multilevel models. Second, we tested to see whether the complete multi-level model fits the data well.

Figure 1

General Model of Math Performance, used in Statistical Modeling, with Demographic Control Variables (Gender and Mother's Education)

