

The Milwaukee Mathematics Partnership:
A path model for evaluating teacher and student effects

by

Carl E. Hanssen, Ph.D.
External Evaluator
Milwaukee Mathematics Partnership

Paper presented at the MSP Evaluation Summit II
Minneapolis, MN, October 4-5, 2006.

Introduction

Alternative statistical models have been proposed for evaluating the effects of the Milwaukee Mathematics Partnership (MMP) on teachers and students. These models have been articulated in the form of hierarchical linear models that incorporate various combinations of student, teacher, classroom, and school-level variables. For example, one HLM approach would nest students within teachers, within schools to estimate the effect of MMP activities on student achievement.

Each attempt to estimate the impact of the MMP on teachers and students, however, share a common, underlying hypothesis: MMP activities, e.g., teacher professional development are impacting teacher mathematical knowledge for teaching (MKT), which is in turn impacting classroom practice, leading to improved student achievement. This paper steps away from the prior HLM analyses and explores these relationships using structural equation modeling techniques.

The path model proposed by this paper hypothesizes relationships between teacher education, teacher experience, professional development hours, mathematical knowledge for teaching (MKT), classroom practice, and student achievement (see Figure 1). We hypothesize that education, experience, and professional development hours are predictors of MKT and classroom practice. We further hypothesize that MKT and classroom practice are predictors of student achievement.

Given the exploratory nature of this work, the purpose of this paper is to present the results from an initial attempt to apply path analysis techniques to the problem of linking MMP activities to teacher performance and student achievement. We expect this

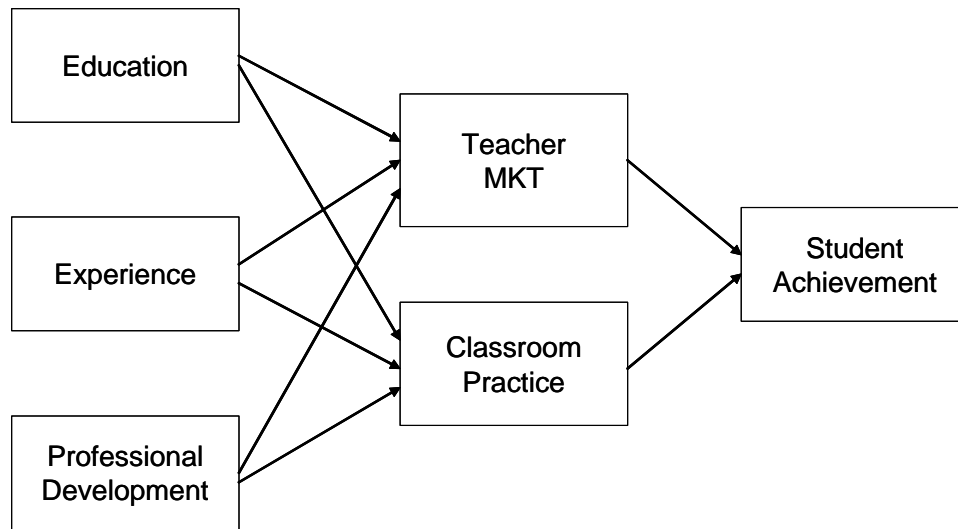


Figure 1. Hypothesized relationships between teacher and student variables

work will be refined in the remaining years of the MMP evaluation. As such, in addition to presenting the results from this work, recommendations for improving the path model, as well as the implications of this work for evaluation are offered.

Method

The following details the methods used for this study. First, the data sources and measures are presented. Next, the final sample is described. Along the way, significant challenges are outlined and discussed.

Measures

All data for this study were collected through the normal course of the MMP evaluation that took place during 2005-2006. That is, we did not collect any new data for this work nor were any measures specifically designed with this analysis in mind. All

data collected used individual teachers as the unit of analysis. The exception to this was student achievement data which was collected at the student level, then aggregated within a teacher. All teachers for this study were from 11 MPS schools that were the focus of in-depth study as part of the MMP evaluation. Multiple evaluation activities took place in each of these schools and math teachers in the schools provided data in a number of areas. Given this, our data collection and subsequent analytical approach centered on linking information about an individual teacher that was obtained from various sources using a unique identifier that existed for each teacher.

Overall, six measures were used based on the evaluation data. They were (1) teacher education level, (2) teacher experience, (3) the number of mathematics professional development hours, (4) teacher MKT, (5) teacher classroom practice beliefs, and (6) student achievement. Each measure is described below.

Teacher *education level* and *teacher experience* were collected from teachers via a survey that used Likert scale items. Responses were assigned a value of 1-5, where a higher value indicated more education or more teaching experience.¹

Professional development hours were tracked by schools and reported to the MMP project team by school personnel. Professional development could have entailed any type of development activity, e.g., seminars, workshops, or individual mentoring, but the developed was to be specifically focused on mathematics. Given that these data were

¹ For education level, 1=bachelors, 2=masters, 3=masters plus, and 4=doctorate (Ed.D. or Ph.D.). For experience, 1=first year teacher, 2=1-3 years experience, 3=4-6 years experience, 4=7-10 years experience, and 5=11+ years experience.

reported by individual schools to the MMP, it is possible that the guidelines for counting professional development hours were interpreted differently by different schools.

Teacher MKT was assessed using a 43-item test. The test contained three subscales in (a) number and operations, (b) algebra, and (c) geometry. Each scale contained 14 or 15 items. Each item was scored as correct (1) or incorrect (0). An overall score was calculated after applying item difficulty parameters² to item responses. The final MKT score used for this study was an IRT score where 0 represents the average score and positive or negative numbers represent scores above and below the mean, respectively.

Classroom practice was evaluated by aggregating responses from 24 survey items that asked teachers about the importance of various classroom practices and the level of implementation of those same practices. For each item, a score of 1-5 was assigned based on the teacher response, where 1 indicated ‘high importance’ or ‘excellent implementation,’ and 5 indicated ‘low importance’ and ‘poor implementation.’

Finally, *student achievement* was based on Fall 2005 WKCE scores for students in grades 3-8. While each student received a scale score, these scale scores are converted to proficiency levels on a scale of 4-1 with 4 being highly proficient. For this study, students who received a score of 4 or 3 were scored as ‘proficient,’ while those with scores of 2 or 1 were scored as ‘not proficient.’ For each set of students associated with a given teacher, a percentage of students who were proficient could be determined. This percentage of students was used as the teacher-level indicator of student achievement.

² Data were normed using a sample of California teachers.

Prior to presenting the analysis, it is important for the reader to understand the difficulties inherent in linking student and teacher-level data for this study. On the teacher side, while data were compiled from multiple sources (paper survey, test, online survey, school reports), it was fairly easy to link these variables together because each teacher has a unique employee ID that is commonly used to identify an individual teacher for payroll and human resources purposes. Because this ID is used to ensure teachers are compensated for their time in providing evaluation data or attending professional development, the ID is easily collected along with the pertinent evaluation data. When data are entered, the ID is captured in the data record along with the survey or test responses.

It was much more difficult, though, to link student achievement data to teachers. At least three major challenges exist. First, student populations in the Milwaukee Public Schools are unstable from the perspective of student mobility. That is, students frequently change schools so that the idea of linking a student test score to a specific teacher may not only be conceptually inappropriate but may also be administratively impossible. Second, the ability to ‘assign’ a student test score to a teacher depends on school-level administrators associating the correct teacher ID to each student. But here the plot thickens because we know, based on experience, that this is (1) a manual process and (2) the ID used for this purpose is not the same ID that is used for payroll and human resources purposes. We also know that this particular ID may change from year to year thus diluting its value as a unique identifier for that teacher. At the same time, there is no reliable Milwaukee Public Schools database that can serve as a ‘key’ for linking the

payroll/HR ID to this other, less stable teacher ID. This entire scenario represented a significant challenge for the evaluation team and it will continue to be a major focus of the MMP evaluation in the coming year.

Sample

The final database for this study contained data for 114 teachers. All were teachers in grades 3 through 8 and all were from the 11 schools that were the focus of the MMP evaluation. The final working database was assembled using MS Access. First, the individual variables were compiled in separate data tables. This was necessary because of the variety of data formats provided. For example, professional development data were provided in an Excel spreadsheet while student achievement data were provided in an SPSS data file. Each variable was then joined to the demographic variable table, which served as the key table for this analysis. Missing data were imputed using the linear trend at point method. To accomplish this, the existing series was regressed on an index variable scaled 1 to n.³ Missing values were replaced with their predicted values. The result of this imputation was a complete data set containing complete information for all teacher subjects. Descriptive statistics are presented in Table 1.

These data show that the typical teacher in our sample has a bachelor degree, has between 4 and 6 years of experience, and participated in 24 hours of mathematics professional development during the 2005-06 school year. Teacher MKT was slightly below average across the sample. In general, teachers believed that that MMP practices

³ SPSS 13.0 for Windows (2004) was used to perform the imputations.

have ‘above average importance’ and that implementation of these practices was ‘above average.’ Finally, on average, just under half of the students in each teacher’s class are rated as proficient or above in mathematics, based on student achievement results.

Table 1

Descriptive Statistics (n=114)

	Study Variable					
	Education	Experience	PD Hours	MKT	Practice	Student Ach.
Mean	1.8	3.9	23.9	-0.19	50.17	47%
SD	.9	1.3	39.6	.79	11.3	15%

Results

Analysis was conducted using LISREL 8.72.⁴ The path model was specified and the covariance matrix (see Table 2) used to estimate model fit. The results showed that model fit was marginal, $X^2(4, n=114) = 9.03, p = .06$; RMSEA = .10; GFI = .97; AGFI = .87. Figure 2 depicts the path model with standard errors.

Table 2

Covariance Matrix

	MKT	Practice	Stdnt. Ach	Education	Experience	PD Hours
MKT	0.62					
Practice	0.41	127.87				
Stdnt. Ach	0.03	0.21	0.02			
Education	0.02	-1.19	0.01	0.80		
Experience	0.04	0.01	0.03	0.43	1.72	
PD Hours	7.30	-34.69	1.63	2.00	4.91	1571.84

Note. Variances are shown on the diagonal.

⁴ Joreskog, K.G., Sorbom, D. (2005). LISREL (Version 8.72) [Computer Software]. Lincolnwood, IL: Scientific Software International.

These results show that the hypothesized model, overall, is not supported by the data used for this study. Only the path between teacher MKT and student achievement is statistically significant yet the results show that MKT accounts for only a small proportion of the variation in student achievement. While this finding supports the MMP position that increased teacher MKT is important for improving student achievement, this findings also underscores the believe that other variables that MMP can not influence, such as SES, may play a more important role in helping predict student achievement.

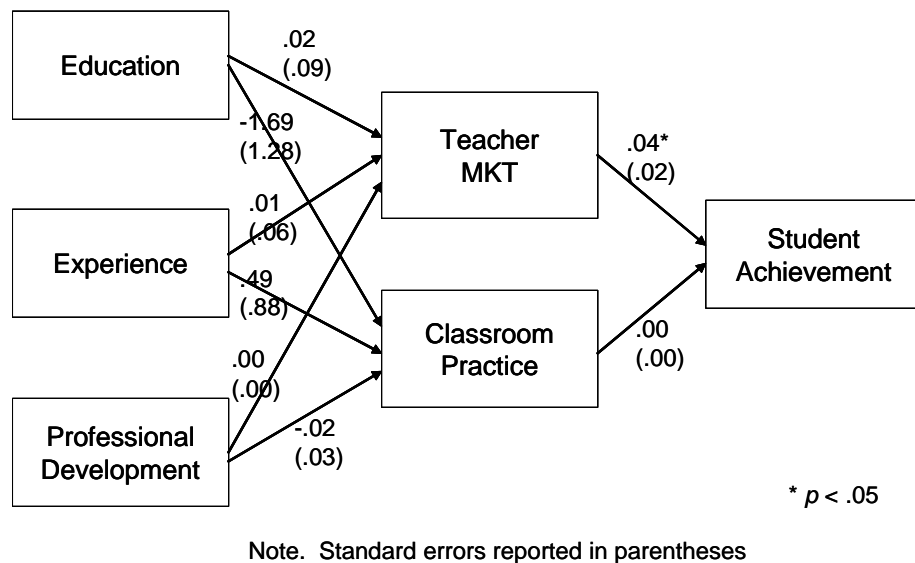


Figure 2. Path model with coefficients and standard errors

Discussion

Given these results, and the exploratory nature of this work, we can at least be encouraged that a model fit solution was obtained. This in itself suggests that there are potential relationships that exist between these variables (i.e., there is some face validity

to the proposed model) but that adjustments to model specification and developing improved measures are needed to develop further evidence. Each of these issues is explored below.

Model Specification

The proposed model may be enhanced in several ways. First, it would seem to make sense that the true value of teacher MKT is in improving classroom practice. Thus, a path might be drawn from Teacher MKT to Classroom Practice, while at the same time eliminating the path directly from Teacher MKT to student achievement.⁵

Second, the paths from teacher experience and education in fact may not be related to MKT. Though common logic suggests that more experienced and better educated teachers would have higher MKT, this in fact may not be the case. We know, anecdotally, that teachers with many years of experience may be more resistant to adopting new ideas about mathematics content and instruction than newer teachers who have less well-formed biases. So it may be that no relationship exists between these variables or at least an inverse relationship might exist (though that would not prevent observing a statistically significant path coefficient if a strong relationship existed).

Third, there may be value in adding components to the model that enrich the proposed theory and that might provide additional explanatory value. For example, recent work by the MMP internal evaluation team suggests that the most important predictors of student achievement are in fact (a) prior year student achievement and (b)

⁵ In fact, this is more consistent with the most recent iterations of the MMP evaluation logic model, which places classroom practice as a mediator between MKT and student achievement.

student socio-economic status.⁶ While these findings might be somewhat discouraging because they suggest that the MMP has little opportunity for impacting student achievement, the addition of these variables into the proposed path model might provide greater insight toward understanding what in fact drives student achievement. These three revisions are depicted as part of a new model in Figure 3 below.

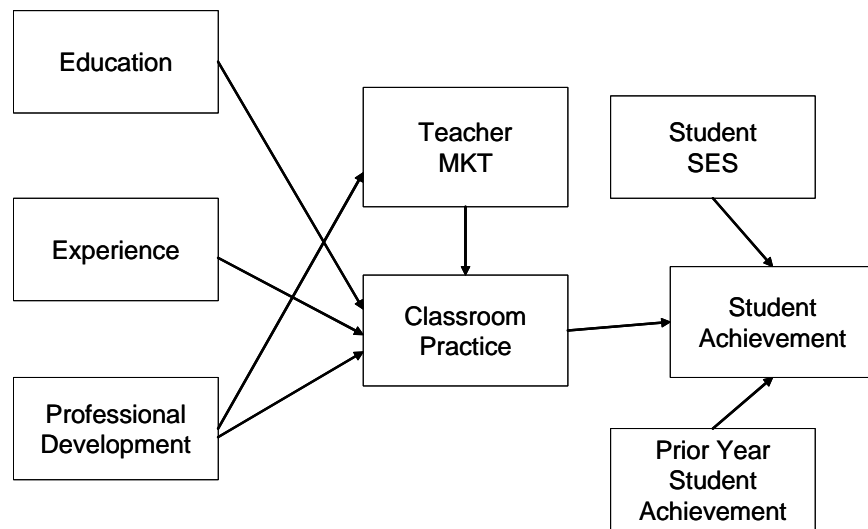


Figure 3. Revised Path Model

Measurement Issues

Several measurement issues emerged during this work. Addressing each has the potential to improve model fit because improved measurement would provide better estimates of actual behavior. First, reducing teacher demographics to categorical data has the effect of reducing variability across the sample and diluting the precision of the

⁶ See Hanssen, C.E. and Walker, C. (2006). Milwaukee Mathematics Partnership: Year 3 Evaluation Report. Milwaukee, WI: University of Wisconsin-Milwaukee.

available data. Alternatives are to collect these data at the interval level (e.g., how many years have you been teaching?) or from existing HR databases. Similarly, education level may be measured with too little precision. A better indication of educational influence on classroom practice might be ‘time since latest degree.’ The rationale here being that individuals that have completed a degree more recently (e.g., within the last few years as opposed to 10 years ago) may have a greater likelihood of being exposed to progressive teaching methods that are more closely linked to proposed classroom practice principles.

Classroom practice measures could also be improved. The current measures rely on teacher perceptions as surrogate indicators for what the teacher is actually doing in the classroom. Though ideally, classroom observations of hundreds of teachers by experienced observers, using a tested protocol would no doubt improve these data, that option is impractical. In light of this, better estimates of the degree to which teachers are applying MMP ideas and concepts in the classroom are needed. Another alternative would be to seek 270 degree or 360 degree feedback about teacher classroom practice. This too, though, may be impractical while facing the added challenge of appearing evaluative of teacher performance.

Lastly, the timing of student achievement data collection is problematic. To illustrate, the data used for this study were from the Fall 2005 administration of the Wisconsin standardized test. These data were linked to teacher data that were collected in Spring 2006 and ostensibly reflective of teacher behavior, attitudes, and knowledge developed during the 2005-2006 school year. The primary question that arises from this

scenario is to what extent is current teacher behavior useful for predicting past (6 months ago) student achievement. Logically it should not be. Rather, we would prefer that the data collected on teacher behavior this year (Spring 2006) be linked to student achievement that will be measured in Fall 2006. And, in fact, in Summer 2007 after we receive Fall 2006 student achievement data, we will be in a position to test this assumption using this approach.

The saving grace in the approach applied here is that because we know that prior year student achievement is highly predictive of current year student achievement, we expect that Fall 2005 achievement data can serve as a reasonable estimate of Fall 2006 achievement. This enables us to test the approach and evaluate its potential given the assumption that many of the issues outlined above can be reasonably addressed.

Implications for Evaluation

This exploratory study, though unsuccessful in providing powerful evidence of links between teacher behavior and student achievement, may prove useful for guiding evaluation activities given there is still interest in using SEM techniques to examine these relationships. Two significant issues emerge—first is what to do in terms of executing the evaluation and second is how to use the information that is collected and the results of any analysis.

First, what to do? This study provides evidence the SEM is a viable method for examining relationships between teacher behavior and student achievement. What is critically missing from this approach, however, are measures that provide valid and

reliable estimates of teacher behavior, especially as they relate to classroom practice. If we believe that classroom practice is the critical factor in driving student achievement (notwithstanding all of the environmental factors), then, as evaluators, we must develop valid and reliable measures of classroom practice. These measure, in turn, need to be constructed so that data can be collected from a large sample of teachers (i.e., reliance on classroom observations is not practical for this purpose) in order to develop a database of sufficient size. An efficient, workable, cost-effective solution to this issue is not readily apparent but the evaluation team will continue to work on this in the coming year.

The other critical action item is to crack the code that will allow us to reliably link student achievement data to teachers. The first step in this process is to work with school district personnel and consult with them on improving their data warehouse so that necessary data are collected, entered properly, and then easily extracted for analysis. The second step is to gain support from school-level personnel who are responsible for providing reliable data to the district (e.g., associating students with teachers). The underlying challenge in this process, however, is that obtaining accurate data depends on multiple individuals at multiple organizational levels touching each piece of data. Like the ‘telephone game,’ each time the message changes hands, it potentially changes. Regardless, the evaluation team will continue to work with district and school personnel to focus on establishing a reliable link between teacher and student information.

Finally, if measurement and data issues can be resolved, the question of primary interest is what to do with this information? From the perspective of evaluation, there are really two purposes to all this work—(1) to document what has occurred in the project so

that others can learn from the experience and (2) to help the MMP project focus its work in areas that are potentially most impactful. The first purpose is clearly summative and for the purposes of this discussion, not as important as the second. The second purpose is formative and this is where evaluation can provide tremendous value. For example, if strong links were found between participation in professional development, classroom practice, and student achievement, then the project would have a strong ‘business case’ for promoting professional development activities. Furthermore, professional development data could then be further analyzed to understand what particular events were more or less impactful. At the same time, classroom practice data could be scrutinized to understand which particular practices appear most directly linked to student performance. This process of ‘peeling the onion’ is laborious, which is why having first having evidence of global relationships can help focus these efforts saving time and resources. The bottom line of this argument, of course, is to use evaluation to help focus the project (and the project’s limited resources) on those activities that have the best chance of improving desired results.

While the results of this study in themselves do not provide much useful evidence about the relationships between teacher behavior and student results, the application of this method is promising. To realize this promise, however, improved model specification and improved measurement must be developed. If these tasks can be accomplished, then the results of this analytical approach may become useful for documenting project accomplishments and providing initial guidance on where and how to focus project activities.