Measuring the Effect of the Milwaukee Mathematics Partnership on Student Achievement

Cindy M. Walker

Jacqueline Gosz

DeAnn Huinker

University of Wisconsin - Milwaukee

Determining the effect of the Milwaukee Mathematics Partnership (MMP) on student achievement in the Milwaukee Public Schools district (MPS) is no easy feat. Within a school, it is ultimately the classroom teacher that has the most direct role in impacting student achievement in mathematics. The MMP aims to influence what occurs at this fundamental level via direct interaction between the Math Teaching Specialists (MTS) at MPS and the Math Teacher Leaders (MTLs) of each school. The end result of this interaction, that may ultimately have a direct impact on increasing student achievement in mathematics, is for the MTLs, in turn, to bring back what they have learned from the MTS and offer support and leadership to others at their school. This is to be accomplished through interactions between the MTL and their school's Learning Team (LT), which, at the very least, is composed of the MTL, the school's Literacy Coach (LC), and the school's principal; as well as with interaction between the MTL and other math teachers in the building. It should be noted that LTs were in place before the onset of the MMP and were designed to serve in a leadership capacity in the school whose main goal was to increase student achievement in reading, writing, and mathematics.

This model of creating school leaders that are to serve as an impetus of systematic change is quite common in educational reform efforts. However, attempting to measure the success of educational reform efforts such as these, in terms of their impact on achievement, is quite difficult because there are many uncontrollable variables at the school level that may affect the desired outcome of increasing student achievement. For example, even the best efforts of an MTL in the MMP will not result in an increase in student achievement if other math teachers in their school are not amenable to change or are not supportive of the efforts of the MTL. Similarly, the efforts of the MTLs may be hindered if others at the school level who have the potential to positively affect student achievement in mathematics (i.e., members of the Learning

Team) are not supportive of reforming the mathematics curriculum. Furthermore, there is no guarantee that the MTLs are actually offering leadership and support to others at their school, especially given the fact that they also have their regular classroom responsibilities to contend with.

In addition to considering extraneous variables, any methodical approach to studying the impact of a large scale educational reform effort on increasing student achievement must decide upon a systematic and plausible approach for assessing accountability. Traditionally most state accountability systems have been status based, relying largely upon cohort-to-cohort comparisons to assess improvement in academic achievement from one year to the next, and thus, school and/or teacher effectiveness. These methods involve comparing unadjusted average student performance on standardized tests across successive cohorts (McCaffrety et al., 2003). For example, the average achievement scores of fourth grade students in year one would be compared to the average achievement scores of fourth grade students in year two in order to ascertain if there was an improvement in academic attainment. However using this approach to determine the effectiveness of a large-scale educational reform effort poses inherent shortcoming because it is based on a nonequivalent-group case study design (Stevens, 2004). Cohort-to-cohort accountability systems do not take into consideration pre-existing differences between successive classes or among schools, either at the individual or institutional level. They do not account for students' incoming knowledge or background factors that are certain to impact performance on achievement tests.

An alternative to using status based accountability systems is implementing value-added assessments to evaluate the effectiveness of a large scale educational reform effort on student achievement. Value-added assessment (VAA) models consider the growth of students from one year to the next and, by using statistical analyses of repeated measures, estimates the unique contribution of each school to this growth, above and beyond prior learning and non-educational factors such as demographics. Tekwe, et al. (2004) define VAA as

> "a term used to label methods of assessment of school/teacher performance that measure the knowledge gained by individual students from one year to the next and then use that measure as a basis for a performance assessment system. It can be used more generally to refer to any method that adjusts for a valid measure of incoming knowledge or ability."

Part of the evaluation of the Milwaukee Mathematics Partnership (MMP) focused on trying to capture some of the school level variability in variables that are not under the control of the MMP, with the thought that any increase in student achievement in mathematics is ultimately affected by differences in how the model of professional development is being implemented in each of the respective schools. The primary purpose of this paper is to determine whether reported differences in these variables are related to differences in student achievement gains in mathematics. For example, is there a greater increase in student achievement in mathematics in schools that report discussing mathematics more often or for schools with a collaborative and supportive LT? A secondary purpose of this paper is to compare and contrast the two aforementioned accountability systems, status based and value-added, in terms of assessing the effectiveness of a large-scale educational reform such as the Milwaukee Mathematics Partnership (MMP). These two methods will be compared in terms of outcomes and viability of application. Finally issues specific to the MMP that have been encountered in attempting to implement a system of accountability will be addressed. These issues are vitally important because they threaten the validity of any conclusions that may be drawn in terms of the impact of the MMP on increasing student achievement in mathematics.

Methodology

Participants

A total of 143 elementary or middle schools participated in the efforts of the MMP during the first year of the grant. However, only 92 schools had an MTL in attendance at the final district-wide MTL meeting that year at which data pertaining to how active schools were in the efforts of the MMP in the first year of the grant was collected. Therefore only those 92 schools could be considered in this study. For the value-added accountability system students could only be utilized if they had achievement test scores for both the 2003-04 and 2004-05 academic years. Under this system students that either repeated a grade or skipped a grade were eliminated from the analysis to ensure that difference scores could be interpreted in the same way for all students. Moreover, students that transferred to a new school in the 2004-05 academic year were also eliminated because it was unclear whether their growth should be attributed to the school they attended in the 2003-04 academic year, which is the school for which MTL data existed, or the school they attended in the 2004-05 academic year. This resulted in a total of 14,259 students in grades 3 through 8 that had were considered in the value-added accountability system. For the status-based accountability system the number of students varied depending on the year considered. In the 2003-04 academic year 27,117 students were utilized and in the 2004-05 academic year 27,079 students were considered. It should be noted that under the status-based accountability system students are utilized regardless of whether they had test scores both years, transferred schools, or repeated or skipped a grade.

Instrumentation

In the first year of the grant a short self-report instrument was designed, based on the goals and objectives of the MMP and administered to all MTLs in the district that were in attendance at the final MTL meeting of the academic year which was held in June of 2004 This survey included items related to the following variables that were believed might differ between participating schools:

1. How often MTLs meet with their Math Specialist and others in their building (i.e., their Learning Team, other math teachers, Literacy Coach, and principal);
2. How satisfied MTLs feel with the amount of time they have to discuss issues centered around mathematics with their Math Specialist and others in their building (i.e., their Learning Team, other math teachers, Literacy Coach, and principal);
3. How supported MTLs feel in their role as a MTL by their Math Specialist and others in their building (i.e., their Learning Team, other math teachers, Literacy Coach, and principal);
4. The MTLs perceptions of the functioning of school's Learning Team in terms of the general climate of the meetings as well as the overall cohesion in accomplishing goals related to their mathematics curriculum.

Actual survey items can be found in Appendix A. MTL responses to this survey were considered to be indicators of how well schools participated in the efforts of the MMP in the first year of the grant and were used as independent variables in the analyses. In addition to the self-report survey data, several other tracking variables that were obtained in the first year of the grant were considered as independent variables. This included the number of district-wide meetings that were attended by the MTL, whether a vision statement was turned in by the MTL, and whether an action plan was turned in by the MTL. A school's vision statement is a concise statement that portrays their vision of what students should be learning in mathematics, as well as how they should be learning it. An action plan is a written request for funding that describes how such funding will be used to support mathematics related work of the learning team and/or the professional development of teachers in the school to achieve that vision.

For all analyses the dependent variable consisted of mathematics scale scores obtained from the November administration of either the Terra Nova or the Wisconsin Knowledge and Concepts Examination (WKCE) in 2003-04 and 2004-05. The WKCE is a state mandated achievement test that schools are required to administer in grades 4, 8, and 10. MPS administers the Terra Nova in the off-years (i.e. grades 3, and 5-7). Both achievement tests are published by CTB-McGraw Hill and research conducted by the publisher has determined that mathematics scale scores obtained from the Terra Nova are highly correlated to those obtained form the WKCE. This suggests that these two tests define and measure the construct of student achievement in mathematics very similarly. Therefore, it is appropriate to directly compare the results of the WKCE and the Terra Nova.

For the value-added accountability system difference scores were calculated for students by subtracting their 2003-2004 mathematics achievement scale score from their 2004-2005 scale score. Then, because the expected difference score for each student is not zero nor is it linear, a weighted average of individual difference scores was calculated to obtain an overall measure of growth for each school. This was done by utilizing the following formula:

$$\frac{\sum_{i-I}^{j} n_{(j-i)}\overline{D}_{(j-i)}}{N} \qquad (1)$$

Where

I = the lowest tested grade level served by a particular school

j = the highest grade level served by a particular school

$n_{(j - i)}$ = the number of students in a particular school who had scale scores for both grades *i* and *j*

$\overline{D}_{(j-i)}$ = the average of individual difference scores for students from grade *j* to grade *i*

and N = total number of students used to calculate the overall growth

Table 1 provides the descriptive statistics for the difference scores obtained for the student participants considered in this study under the value-added accountability system.

Table 1

Descriptive Statistics of Difference Scores Obtained from Value-added Accountability System

$\overline{D}$ (2004/5 –2003/4)	Schools	N	Mean	SD	Min	Max
All cohorts	92	14,259	16.27	28.30	-200	206
4th Grade - 3rd Grade	76	2,960	23.29	33.30	-200	183
5th Grade - 4th Grade	76	3,252	8.37	26.55	-172	183
6th Grade - 5th Grade	76	1,376	11.69	26.38	-106	174
7th Grade - 6th Grade	53	3,313	15.17	26.01	-194	206
8th Grade - 7th Grade	38	3,358	20.69	25.58	-140	183

For the status based, or cohort-to-cohort, accountability system the average scale score for each grade was calculated for each school for both the 2003-04 and the 2004-05 school years. The difference in these average scale scores for the two school years was then calculated for each grade. For example, for a particular school the difference between the average scale score obtained from their 3rd graders in the 2003-04 academic year and that obtained from their 3rd graders in the 2004-05 academic years was considered to represent growth in student achievement for third grade students at that school. Then, because the expected difference for each grade level is zero, the average difference across all grade levels was calculated to obtain an overall measure of growth for each school. Table 2 provides the descriptive statistics for the difference scores obtained for the student participants considered in this study under the status-based accountability system.

Table 2

Descriptive Statistics of Difference Scores Obtained from Status-based Accountability System

Cohort to Cohort (2004/5 –2003/4)	Schools	N_{2003}	N_{2004}	Mean	SD	Min	Max
All cohorts	92	27,117	27,079	-2.55	7.39	-20.64	14.63
3rd Grade	76	4,166	3,977	-3.05	14.77	-49.63	41.59
4th Grade	76	4,598	4,147	-5.10	10.77	-28.76	32.09
5th Grade	76	4,270	4,287	-5.59	11.52	-45.98	21.56
6th Grade	53	5,047	4,721	0.37	9.83	-40.37	22.40
7th Grade	38	4,577	4,763	-3.66	12.03	-38.94	25.06
8th Grade	34	4,457	4,490	7.91	11.79	-24.06	30.88

Analyses

Regression analyses and ANOVAs were conducted to determine which, if any, of the school level variables helped to explain differences in student achievement gains in mathematics. To ensure that the variables created were meaningful, the results obtained from the analyses were meaningful and multicollinearity did not exist, exploratory analyses were conducted on the MTLs responses to survey. Specifically, cross-tabulations and correlations among various pairs of items were examined. Based on these preliminary analyses the following four evaluation questions were developed to guide the analyses:

1. Does the degree in which schools' learning teams embrace the vision of the MMP influence student growth in mathematics achievement test scores?
2. Does the level of satisfaction with time spent discussing mathematics, as reported by MTLs, and frequency of MTL attendance at district-wide MMP meetings influence student growth in mathematics achievement test scores?
3. Is there a difference in student growth in mathematics for schools for which MTLs report feeling a greater ability to accomplish goals regarding mathematics?
4. Is there a difference in student growth in mathematics for schools in which MTLs feel more supported by other math teachers in their school?

To answer the first research question a stepwise regression analysis was conducted for each of the accountability systems considered, using overall school growth in mathematics as the dependent variable. A significance value of 0.15 had to be met before an independent variable was allowed to enter the regression model. Five independent variables, thought to represent the priority that schools' LTs placed on the vision of the MMP, were considered. The first independent variable was created by coding the frequency of LT meeting such that a score of six represented monthly meetings, a score of 12 represented every other week, and a score of 24 represented weekly meetings. These scores were then multiplied by the scores obtained the survey item that asked MTLs how often they had useful discussions about what should be done to help students succeed in math. For this survey item a response of "always" was coded as a 1, a score of "most of the time" was coded as 0.75, a score of "some of the time" was coded as 0.25, and a score of "never" was coded as 0. The second independent variable was created to represent the climate of the learning team. This was accomplished by averaging the item level responses to the first four items in question 4 which can be found in Appendix A. The reliability of this scale was 0.81. The third independent variable was created to represent the level of LT collaboration. This was accomplished by averaging the item level responses to items 5 - 7 in question 4 which can be found in Appendix A. The reliability of this scale was 0.88. Two dichotomous variables were included in the regression analysis. These included whether a school had submitted a vision plan and whether they had submitted an action plan. These plans were supposed to be created by the entire LT. However, if the climate or collaboration level of the LT was not ideal it is conceivable that this did not happen. Therefore, four interaction terms were included in the regression analysis which crossed the two dichotomous variables with the LT climate variable and the LT collaboration variable.

To answer the second evaluation question a stepwise regression analysis was conducted for each of the accountability systems considered, once again using overall school growth in mathematics as the dependent variable and a significance level of 0.15 to enter. The independent variables in this regression model included the number of district-wide MMP meetings attended by the MTL, how satisfied the MTL was with the amount of time spent discussing mathematics with others and the interaction between these two main effect variables. The variable representing the amount of time spent discussing mathematics with others was created by averaging all of the item level responses to question two which can be found in Appendix A. The reliability of this scale was 0.83.

To answer the third evaluation question a one-way ANOVA with three levels was conducted using overall school growth in mathematics as the dependent variable and the response to the last item of question 4 of the survey as the grouping variable. Due to the low number of MTLs (n = 3) that reported never feeling able to accomplish their goals related to mathematics the response categories of “never” and “sometimes” were collapsed. Similarly, to answer the fourth research question a one-way ANOVA with three levels was conducted using overall school growth in mathematics as the dependent variable and the response to the second item of question 3 of the survey as the grouping variable.

Results

Given the fact that all of the analyses were conducted based on data that only represented approximately six months of grant activity it is not surprising that very little statistically significant results were obtained. What follows is a brief description of the results obtained for each of the evaluation questions.

Evaluation Question #1

For the value-added accountability system none of the independent variables were significant enough to be entered into the regression model. However for the status-based accountability system the variable representing LT collaboration was found to be a significant predictor of student growth in mathematics ($F_{1,88}$ = 5.06, p=.027, β = -2.23, R^2 = .05). The negative regression weight suggests that as the level of LT collaboration increases the overall school level growth in mathematics achievement decreases which is an obviously incongruent finding to what would be expected theoretically.

Evaluation Question #2

For both the value-added accountability system and the status-based accountability system MTL attendance at district-wide MMP meetings was found to be a statistically significant predictor of overall school level growth in mathematics achievement. For the value-added accountability system $F_{1,88}$ = 4.17, p=.044, β = .98, R^2 = .05. For the status-based accountability system $F_{1,90}$ = 6.37, p=.011, β = 1.31, R^2 = .07. It is interesting that the results are the same for both accountability systems and unclear why the results appear stronger for the status-based accountability system since this is not a theoretically appropriate accountability system. More importantly it is encouraging that attendance at the district-wide MMP meetings seems to play a role in increasing student achievement in mathematics in the district, especially given the fact that the results pertain to only approximately six months of grant activity.

Evaluation Questions #3 and #4

No statistically significant differences were found for either of these evaluation questions for either of the accountability systems considered. This may be due to the fact that, although all of the survey questions used as independent variables were self-report, the survey questions that were utilized to form groups for these evaluation questions pertained to MTLs feelings, as opposed to more definitive concepts.

Discussion

The primary purpose of this part of the evaluation of the MMP was to determine if reported differences related to how well schools participated in the efforts of the MMP during the first year of the grant were related to differences in student achievement gains in mathematics. A secondary purpose was to compare and contrast the outcomes and viability of status based and value-added accountability models in terms of assessing the effectiveness of a large-scale educational reform such as the Milwaukee Mathematics Partnership (MMP). Given the fact that the MTL survey responses, used as independent variables for the analyses, were obtained in June, only six months after the inception of grant activities, it is not surprising that the analyses did not detect many relationships. On the other hand, it is encouraging that a small positive effect on overall student achievement in mathematics was found in terms of MTL attendance at district-wide MMP meetings.

It is interesting that so few differences were found between the two different accountability systems considered in this evaluation. However, the differences that were found highlight the major flaws of using a status-based growth model, namely it utilizes different cohorts of students to ascertain a measure of change in achievement. While the difference scores obtained from the value-added model were all positive, the difference scores obtained from the status based model were primarily negative (see Tables 1 and 2). Moreover, the magnitudes of the difference scores obtained from the value-added model were all larger, when compared to the difference scores obtained from the status-based model. This also was observed in the results of the regression analysis conducted pertaining to the first evaluation question. Specifically, although the collaboration of the LT was found to be a significant predictor of overall school level growth in mathematics achievement, it was found to be negatively related to growth. This would seem to be an anomaly and demonstrates the problem of not taking into consideration students' prior achievement when measuring growth. A likely explanation for the anomalous result is a large number of the schools that had a more collaborative learning team had greater challenges to deal with at the onset of the second year of the grant.

Currently value-added assessment (VAA) growth models are being used across the nation to identify effective schools because they attempt to show the impact an individual school may have on students' rate of learning. The use of longitudinal data to estimate a school's effect on student learning is a unique feature of value-added accountability systems and is likely to prove valuable in assessing the effectiveness of large-scale educational reform on student achievement. This is due to the fact that VAA compares students' progress to his or her past academic achievements rather than to a generic cohort of students and thus, offers a picture of the rate at which a student is learning. Furthermore, through the tracking of individual student progress from one grade to the next, VAA models provide a control over the individual differences between students because students serve as their own controls.

While value added assessments provide many advantages over status-based methods of accountability, there are many issues surrounding the use of VAA models that need to be addressed, particularly given the serious ramifications that may result from erroneous results related to the effectiveness of large-scale education reform. The main disadvantages of VAA stem purely from a practical standpoint. VAA models require an intensive database of quality longitudinal data containing achievement scores, student characteristic variables, and a means of linking students to educational reform activity over the years. This requires reliable and valid means of quantifying educational reform endeavors and outcomes, which is not easy and may not have been accomplished during the first data collection efforts pertaining to the evaluation of the MMP. It is unclear if what the MTLs reported was actually representative of what occurred in their schools because there was no evidence to support their responses. The data collection activities that occurred in the second year of the grant attempted to rectify this by surveying all LT members and teachers of mathematics to try and cross-validate responses.

There are also a number of theoretical concerns that must be considered when utilizing VAA growth models in the evaluation of large-scale educational reform on student achievement. Such issues include model specification, the viability of using the model to claim causal effects of the reform on student learning, the inclusion of covariates and possible confounding variables, the use of achievement test data as a primary outcome measure, and the effect of having missing data that is most likely not missing at random. The final theoretical concern is crucial because the longitudinal nature of VAA models is highly subject to missing data and it is likely that such data is not missing at random. If the missing data *are* missing completely at random, that is, there are no discernable features that characterize the students' who have missed a test from those who have not; then there is little threat to the estimated effect of large-scale educational reform. However, students who do miss tests often are lower achieving and have certain background characteristics that distinguish them from the population of students who take the tests. As such, this data cannot be ignored because it introduces bias into the estimates of school-level effects.

The estimation of large-scale educational reform effects obtained from VAA modeling of may also be impacted by the time at which the measure of student achievement used in the model is administered. It is unclear how to model growth in mathematics achievement so that a student's growth is accurately distributed among the schools responsible for that growth, especially when trying to explain that growth by school level variables that are collected at a different point in time. This has been a major issue in trying to model how variability at the school level, which is related to the efforts of MMP, is having an impact on increasing student achievement in mathematics. While the efforts of the MMP start in September and end in August, the school year runs from September to June, students are tested in November, and data is collected from teachers, learning team members, and MTLs in June.

In addition, when growth is used as a dependent variable there is an inherent implication that the scale of measurement is constant across years. This requires tests from one year to the next to be vertically scaled. In doing so, the assumption is made that each test is measuring a common trait so that the scales have identical meanings. However, as students progress from one grade to the next, the content of their curriculum progresses in content and difficulty. Therefore, tests designed for different grade levels may not allow for a measure of true growth. Standardized achievement tests are also subject to measurement error and the amount of error varies depending on a student's level of achievement, with the greatest amount of variability at the extremes of the ability distribution. This fact can lead to a distortion of estimated reform effects, most notably when students of similar characteristics are clustered within a particular school.

Another problem that has been encountered in trying to determine the relationship between the efforts of the MMP and student growth in mathematics achievement is that currently only school level variables can be utilized. This is true not only for the present component of the evaluation but will continue to be an issue in the upcoming years because of issues associated with teacher confidentiality and lack of participation in data collection efforts. There is currently no way to link teacher responses to the students in their classrooms. Moreover, the response rates for the data collection efforts that took place at the culmination of the second year of the grant which were aimed at all school employees that could possibly impact student achievement in mathematics were extremely low for some schools. Even if there were a way to link teacher responses to students it is likely that modeling classroom effects would be statistically biased due to the magnitude of missing responses.

References

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating Value-Added Models for Teacher Accountability*. Santa Monica, CA: RAND Corporation.

Stevens, J. (2004, October). *The study of school effectiveness as a problem in research design*. Paper presented at the conference of Value-Added Modeling: Issues with Theory and Application, College Park, MD.

Tekwe, C. D. et al. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics*, 29, 11-36.

Appendix A
MTL Item Level Response to Survey

Question 1 **How often do you meet with the following:**	**Monthly**	**Every other week**	**Weekly**
your school's learning team. (n = 95)	29	28	38
other math teachers in your school. (n = 66)	33	10	23
your school's literacy coach. (n = 78)	10	18	50
your school's principal. (n = 82)	7	26	49
your school's district math specialist. (n = 44)	41	3	0

Question 2 **I'm satisfied with the amount of time I have to discuss math issues with:**	**Always**	**Most of the time**	**Some of the time**	**Never**
my school's learning team. (n = 94)	16	36	32	10
other math teachers in my school. (n = 91)	4	29	46	12
my school's literacy coach. (n = 87)	19	32	29	7
my school's principal. (n = 94)	18	38	34	4
my school's district math specialist. (n = 91)	18	31	35	7

Question 3 **I feel supported in my role as Math Teacher Leader by:**	**Strongly Agree**	**Agree**	**Disagree**	**Strongly Disagree**
my school's learning team. (n = 97)	36	52	8	
other math teachers in my school. (n = 93)	21	55	17	

my school's literacy coach. (n = 93)	45	40	8	
my school's principal. (n = 96)	45	45	6	
my school's district math specialist. (n = 96)	40	49	7	

Question 4 **During Learning Team meetings:**	**Always**	**Most of the time**	**Some of the time**	**Never**
the general atmosphere and climate is friendly. (n = 97)	54	39	4	
I am an active participant. (n = 97)	54	32	11	
I feel comfortable expressing my opinions. (n = 97)	65	22	9	1
differences among members are respected. (n = 97)	46	43	8	
members share responsibility for making decisions about school wide implementation of the mathematics program. (n = 96)	39	29	23	5
members work together to make decisions about school strategies to improve student achievement in mathematics. (n = 97)	36	36	21	4
there is a sharing of ideas about how to help students succeed in math. (n = 97)	17	41	36	3
there are useful discussions about what should be done to help students succeed in math. (n = 97)	15	43	36	3
I feel I am able to accomplish goals regarding issues surrounding mathematics. (n = 97)	18	39	37	3