2008

The Vermont Mathematics Initiative: Student Achievement from Grade 4 to Grade 10 2000 Through 2006

H. W. Meyers
University of Vermont

Douglas Harris
The Evaluation Center, The Vermont Institutes, Montpelier, VT

Title: The Vermont Mathematics Initiative:
Student Achievement from Grade 4 to Grade 10

The Vermont Mathematics Initiative: Student Achievement from Grade 4 to Grade 10 2000 Through 2006

H. W. Meyers
University of Vermont

Douglas Harris
The Evaluation Center
The Vermont Institutes, Montpelier, VT

Paper presented at the Annual Meeting of the American Educational Research Association
New York, NY
March, 2008

Title: The Vermont Mathematics Initiative:
Student Achievement from Grade 4 to Grade 10

Abstract

The Vermont Mathematics Initiative (VMI) has been under development and implementation since 1999. The program's evaluation documents the recruitment of teachers; collaboration among a university, a state Department of Education, and school districts; and development of a comprehensive content-rich mathematics curriculum for teachers. The initial VMI report documented statistically significant and educationally important gains for students in intervention schools. In this subsequent evaluation, a third data point over eight years of implementation permits the longitudinal study of gains observed over a longer period. Findings from this study indicated additional significant effects for the intervention that were amplified for students who had been taught by teachers in the second or third year of the teacher's enrollment in the VMI program.

Purpose of this paper

The purpose of the paper presentation is to present findings from Year 8 of an eight year study of the process and value added effects of a statewide, content rich, mathematics professional development program known as the Vermont Mathematics Initiative (VMI).

The paper presentation will:

- Describe the purpose, structure, staffing, and content of the instructional program for teachers.
- Describe the logic model that provides the framework for evaluation of the program.
- Examine the methodology for assessing the value added to student learning by teacher preparation in mathematics.
- Present findings from a longitudinal analysis of two cohorts of students who were taught by VMI participants. Each such cohort of students was matched with a cohort of control students. All student achievement is followed from grade 4 to grade 8 and grade 10.
- Discuss issues of reliability of the results and plans for future evaluation.

Outline of the Design of the Evaluation

For the intervention, 233 students who began with VMI teachers in 1999-2000 at grade 4 were tracked forward to grade 8 in 2004 and further to grade 10 in 2006 (spring testing). These students were compared with an equivalent number (256) of students who were in schools matched with the intervention schools at grade 4. Criteria for matching included, continuous enrollment in the district to grade 8, school size, and concentration of low income students (Stuart, 2007).   In the 2005 study, despite the varying levels of intervention as indicated by both school size and number of VMI enrolled teachers, the Grade 8, VMI taught schools outperformed control schools (Meyers, 2005).

Based on the results of the previous year's evaluation, new evaluation questions emerged for the current year. When the intervention group of students was divided according to whether the school had multiple teachers trained in VMI or just one teacher trained in VMI, we found that the results in intervention schools were significantly different. Greater concentrations of VMI trained teachers accounted for higher test scores in schools matched on other variables. This finding suggested that the extended longitudinal analysis should take into account either the concentration of teachers in the school participating in the VMI intervention model or focus on only those students within the schools who were taught by VMI enrolled teachers for at least one year prior to the baseline measure (Meyers, 2005).

Terminology used in this paper: To differentiate among teachers based on their preparation for mathematics teaching, the term VMI Teacher to refer to the grouping of teachers who have participated in the VMI professional development intervention at a time when they would have made at least one year of contribution to a students development in mathematics. Thus, a VMI Teacher refers to a teacher who has taught a group of students during or one year prior to the baseline measurement year (spring 2000). This definition resulted in a subset of 16 teachers and 233 students who had been taught by them for at least one year.

This paper presentation, therefore, focuses on the following research questions:

- Does student achievement differ in subsequent years between groups of students who were taught by teachers who attended and completed the VMI training program in either 1999 or 2000 and a random sample of those who were not taught by VMI teachers?
- Does student achievement differ between and or among groups of students formed by the intervention and control and eligibility for free or reduced lunch with respect to the growth both groups experience over a period of 8 years?
- Does student achievement on different types of mathematics tasks, namely skills, concepts and problem solving differ between and or among groups of students formed by intervention and control and eligibility for free or reduced lunch?
- Is there a relationship between self reported measures of opportunity to learn mathematics at grades 4 and 8 between groups of students who were taught by teachers who attended and completed the VMI training program in either 1999 or 2000 and a random sample of those who were not taught by VMI teachers?

Perspectives

The VMI is a mathematics content-intensive, comprehensive statewide professional development program, focused at the elementary and middle level, whose mission is to significantly improve the teaching and learning of mathematics in the elementary schools of Vermont. The program is partially funded by local district sources and by a state Title IIB Mathematics and Science Partnership grant. In the spring of 2004, the program received national attention from *Education Week* (Galley, 2004, April 7). Measurement of student performance in mathematics has developed from the State's first attempts at portfolio assessment (Meyers and Brewer, 1990) through the development of state networks to support standards based mathematics instruction. Vermont's portfolios have been the focus of early studies of both teacher and student performance. (Stecher & Mitchell, 1995).

In brief, through their VMI experience, teachers build strong mathematics content knowledge, develop their ability to conduct action research about their teaching practices, cultivate leadership skills, and bring all of this acquired knowledge and skill to bear in their classrooms and at the school or district level. The school leadership component of the VMI incorporates the collaboration of the VMI teacher and the school principal in developing their "VMI impact plan" to improve mathematics teaching and learning in the school. This three-year program leads to a M.Ed. degree from the University of Vermont with a specialty in K-8 mathematics teaching. The VMI design also calls for training a cadre of mathematics teacher leaders who in turn will enhance the teaching of other teachers in the school, and ultimately impact the mathematics learning of all children in the school.

VMI employs a Value Added evaluation design. While there is continuing debate about the errors inherent in various methodologies (covariate adjustment, gain score, and multivariate models) McCaffrey and colleagues (McCaffrey, D.F., et al, 2003) conclude that teacher effects greater than zero are likely and that teachers can and do "differentially effect student achievement" (p.113). For VMI, slope and intercept comparisons of cohorts, over time, have been used to estimate the long term effects of short term teacher development programs on student achievement with single group time series designs.

Many teacher development programs have sought an evaluation model that measures teacher effects while controlling for student and school effects (Stuart, 2007). Such models are problematic in most states because yearly student level data with vertical scalar measurement is not available. However, new federal requirements for annual testing in grades three through eight offers the promise of cohort data from assessments that provides coherent measures of standards within states at each grade level. Until such assessment data becomes available, states evaluating the Mathematics and Science Partnership programs funded with Title IIB NCLB resources must approximate evaluation designs that make the best use of available data.

Recommendations for evaluation of mathematics professional development programs from the National Academies of Sciences (2004) include the identification of implementing weak and strong implementing teachers as a strategy to identify the variation in correlated student achievement. (p 114). The prepublication copy of the report also recommended the use of multiple outcome measures and measures of levels of implementation of curriculum as evaluation strategies (p. 2). Raudenbush (2005) recommends a goal for research on intervention programs "to study why an intervention works for some children and not others, or why it fails." A recent RAND report (Le, et al, 2006) recommends measures that better match interventions and, in particular, the use of subscales for "a more refined analysis." (p. 75). Finally, a recent RMC authored report, ***State Mathematics and Science Partnerships Year I Implementation,*** (2004) recommends both outcome and process evaluation designs and measures. The evaluation of the VMI described herein builds on the recommendations of these five sources.

Vermont has had the rare opportunity afforded by an integrated, statewide, individual student data base. The data base, which has been operational since 1999, is linked by a unique student identification number that enables longitudinal study of common student performance measures. The paper described in this proposal illustrates how it is possible to address goals of both outcome and process evaluation by following cohorts of students and their teachers over time.

Methods

The long term evaluation of the VMI (Meyers & Harris, 2004) employs mixed methods of both qualitative and quantitative strategies. Qualitative strategies included:

- measuring teacher knowledge of program content,
- document analysis of teachers' work products, including action research project reports

- review of course materials developed and implemented in institutes and mentoring sessions
- review of recorded interviews and questionnaires with teachers and administrators, and
- direct observation of the summer institutes and mentoring processes that constitute the delivery of mathematics content and its application in the classroom.

Quantitative strategies included:

- a quasi-experimental design wherein Matched groups of students (Stuart, 2007) are compared for intervention effects Matched groups on gender and SES are drawn at random from schools in a statewide data set.
- a longitudinal design wherein individual students are tracked across years at three data points (grades 4, 8 and 10) and compared for both pre and post differences within groups and between the intervention and control schools at each data point. T-tests were employed for scale scores. Chi-square analysis was employed to determine levels of significance for comparisons of groups formed by both intervention and control and free-lunch eligibility with respect to performance levels.

Data Sources

Program outcomes were tracked over time by:

- Utilizing state assessment data gathered from the annual administration of the *New Standards Reference Examination* (NSRE) mathematics tests at grades 4, 8 and 10. The Vermont Assessment for the years 1999 through 2006 was adapted from the NSRE developed at the University of Pittsburg. Questions were added by the state to measure specific skills, concepts and problem solving as well as opportunities to learn mathematics and English language arts as reported by students.
- Isolating *Stanford Achievement Test* (SAT9) items and scales that are extracted from student level NSRE records for each of 2 cohorts of students whose teachers show varying levels of participation in the program.
- Matching student records over time from grade 4 to grade 8 to grade 10 within the Vermont state student level data set.

Limitations

The following limitations apply to the design of the study:

- Sample attrition. Students where selected as a sample of all those taught by VMI Teachers who formed the 1999 and 2000 cohort, teaching grade 4, by selecting all students who were so identified by the teacher's name on the NSRE student records for the year 2000. About 12 percent of these students withdrew from the cohort of students prior to 2006. Student scores and performance levels for 2006 were carried forward from 2004 as a method to balance the attrition in both groups. The same weighting procedure was applied to the control sample. Control students were selected randomly from the pool of grade 4 students on the NSRE 2000 file. Approximately 30 percent of the control sample withdrew. The substitution of the 2004 performance level for the 2006 level enabled mean comparisons that represented the performance levels of both groups adjusted for withdrawal. The resulting sample of students that formed the student VMI and Control groups is described in Table 1.

- Validity of the measures. As a standards-based measurement of skills, concepts and problem solving there has been a tacit assumption that the measures of mathematics in the NSRE (New Standards Reference Examination), including the Stanford Achievement (SAT9) items, were appropriate to gauge some contribution of the VMI program intervention. While an analysis of the match between the content of the VMI courses and the NSRE measures has yet to be done, the emphasis on problem solving and topics such as probability and statistics in the VMI is to some extent reflected in the NSRE measures.
- Reliability of the scoring from one test form to the next with respect to constructed response items which form the basis for some of the performance level measures prevents the use of trend analysis for these measures. The SAT9 scale scores tend to perform in a linear way over time for cohorts of students while the performance levels are subject to performance level adjustment within

years rather than across years. Hence, longitudinal comparisons are not appropriate for these measures.

Table 1
Tabulation of the VMI and Control Samples by Free Lunch Eligibility

			VMI00		
			Controls	VMI	Total
Free Lunch Numeric	Not Eligible	Count	191	181	372
		% within VMI00	74.3%	77.7%	75.9%
		% of Total	39.0%	36.9%	75.9%
	Eligible	Count	66	52	118
		% within VMI00	25.7%	22.3%	24.1%
		% of Total	13.5%	10.6%	24.1%
Total		Count	257	233	490
		% within VMI00	100.0%	100.0%	100.0%
		% of Total	52.4%	47.6%	100.0%

Results

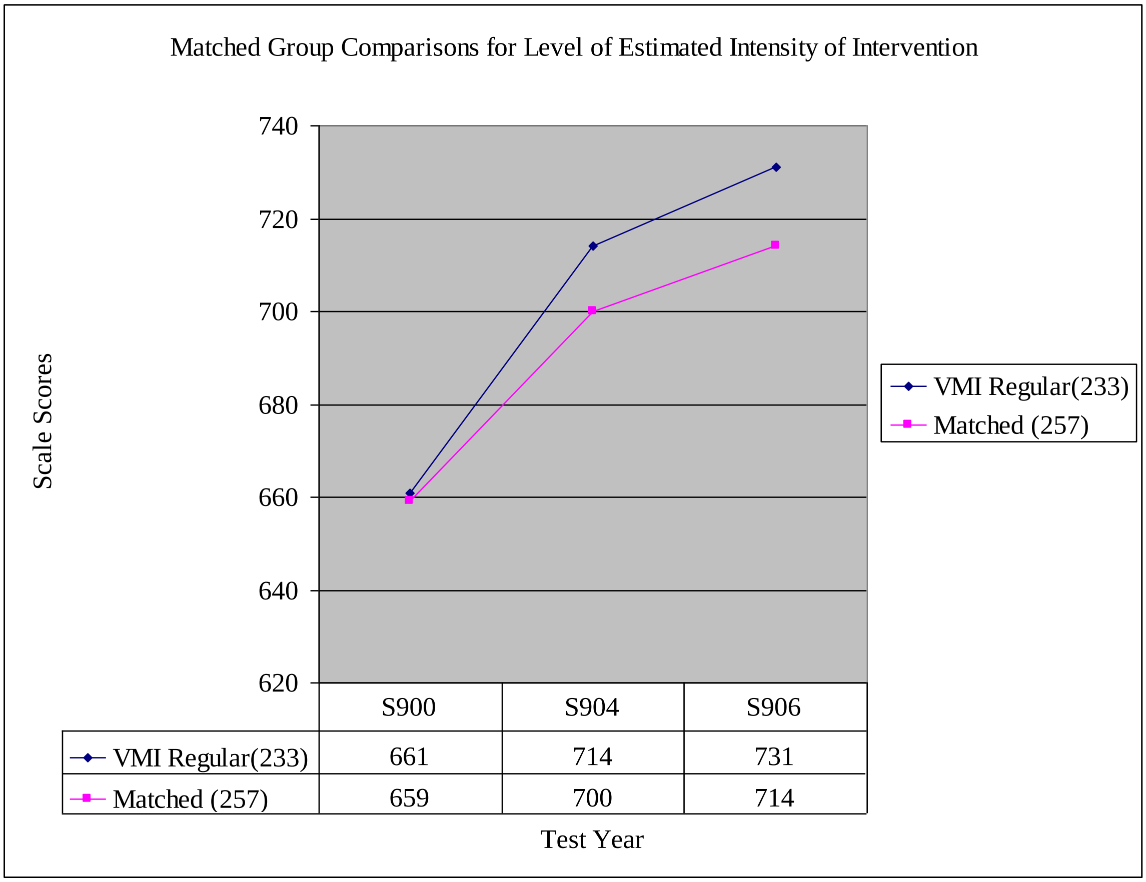
Each of the research questions is addressed below:

- Does student achievement differ in subsequent years between groups of students who were taught by teachers who attended and completed the VMI training program in either 1999 or 2000 and a random sample of those who were not taught by VMI teachers.

In order to evaluate the above question, groups of student scores formed the intervention and by the pairs of scores at each data point were compared. Figure 1 below indicates the results of the longitudinal comparison of the SAT9 scale scores.

Title: The Vermont Mathematics Initiative:
Student Achievement from Grade 4 to Grade 10

Figure 1



	S900	S904	S906
VMI Regular(233)	661	714	731
Matched (257)	659	700	714

The differences between the VMI and Control groups was not significant in 2000 at the baseline, but proved to be significant at 2004, t=-3.07 (equal variances not assumed), df 488, p<.002; and significant again at 2006, t=-2.99 (equal variances not assumed) df 488, p=.003.

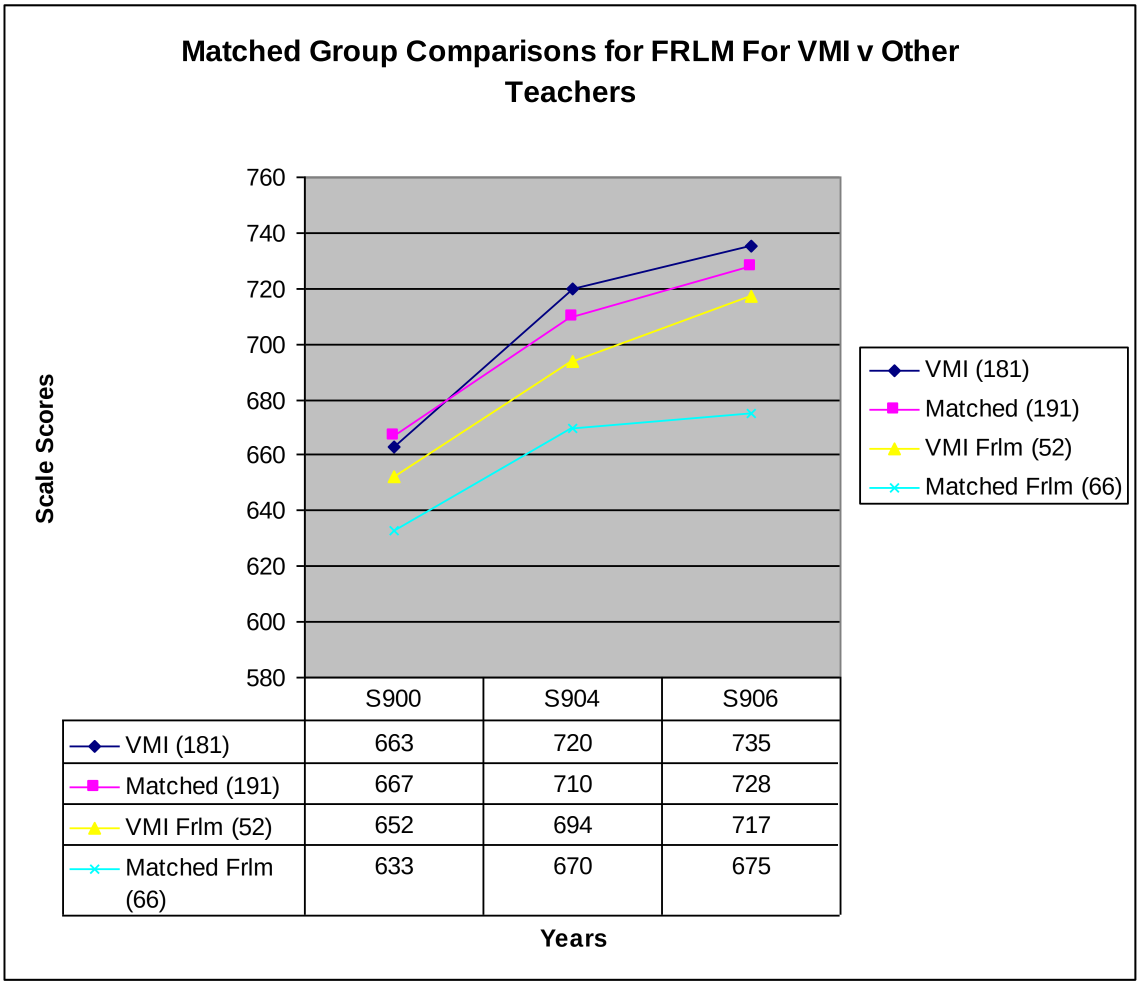
A second comparison to determine within group differences over time, compared the pairs of scores for each group at the subsequent data points of 2004 and 2006. For the VMI group of students the change from the mean of 661 in 2000 to the mean of 714 in 2004 was significant (t=-21.01, df, 232, p<.001). For the comparison of the VMI taught students from 2004 to 2006, the difference was also significant (t=-6.85, df 232, p<.001).

For the Control group of students comparisons between the 2000 and 2004 means was statistically significant ($t=-16.24$, df 256, $p<.001$). In the comparison of 2004 to 2006 the gain of 16 points on the scale score was also significant ($t=-8.01$, df 256, $p<.001$).

- Does student achievement differ between groups of students formed by the intervention and control and eligibility for free or reduced lunch with respect to the growth both groups experience over a period of 8 years?

In order to evaluate the relationship between intervention and family background groups of students formed by both the intervention and family background were compared on the outcome measure of the SAT9 scale scores. Figure 2, below, describes these relationships.

Figure 2



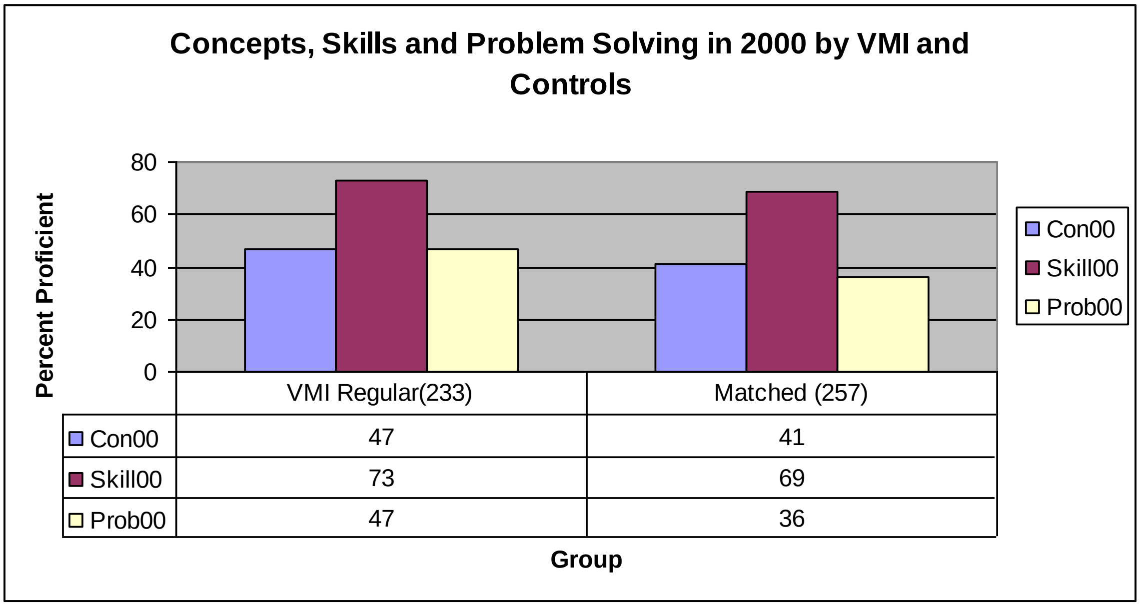
	S900	S904	S906
VMI (181)	663	720	735
Matched (191)	667	710	728
VMI Frlm (52)	652	694	717
Matched Frlm (66)	633	670	675

Point in time comparisons among the four groups was performed for each year with analysis of variance and post-hoc comparisons. Only the Control Frlm (Control eligible for free or reduced lunch) was significantly different from the VMI (non-eligible) and Control (eligible) groups in a Bonferroni post-hoc comparison (p<.001). This pattern of difference was repeated at grade 8 (S904) and, in addition, the Control eligible group was then significantly lower from the VMI eligible group (p<.05). The grade 8 pattern was again repeated at grade 10, but this time the Control eligible group was even lower than the other three groups (p,.001). Inspection of the data points for this group revealed that the leveling effect experienced by the Control eligible group was linked to early withdrawal from school. At grade 10, 30 percent of the cohort of Control eligible students had dropped out of school while only 12 percent of the VMI eligible group had dropped out. These dropout rates compared with 10 percent of the VMI non-eligible and 15 percent of the non-eligible Controls. These numbers compare favorably with the overall state graduation rate of 82 percent since the graduation rate is most closely aligned with the cohort completion rates (Vermont State Department of Education, 2007).

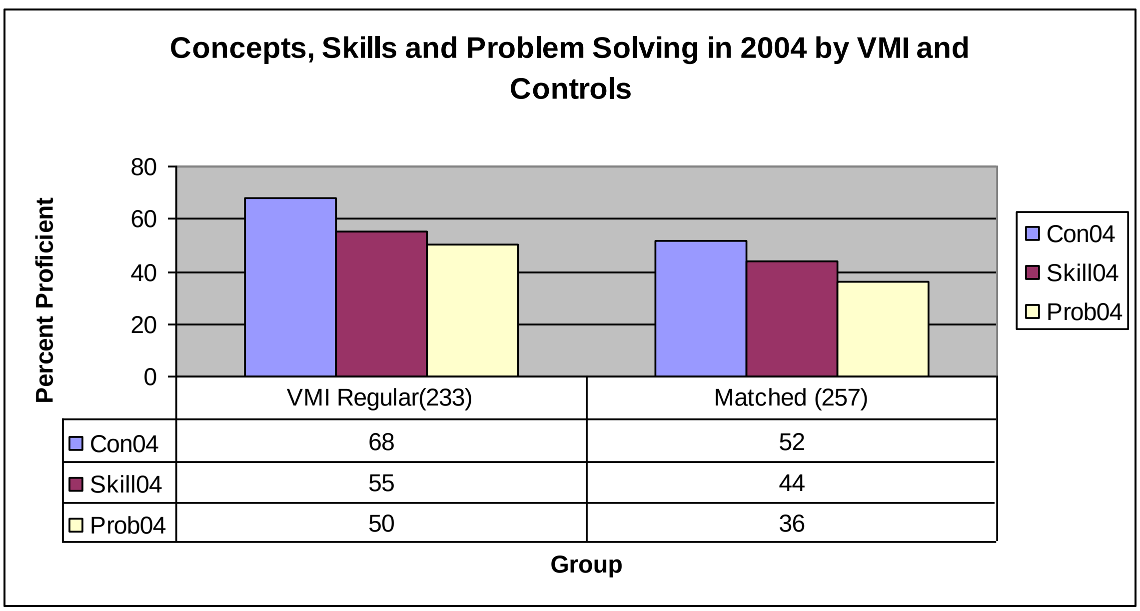
- Does student achievement on different types of mathematics tasks, namely skills, concepts and problem solving differ between groups formed by the VMI intervention and the Control group; and, does performance differ between or among groups of students formed by intervention and control and eligibility for free or reduced lunch?

In order to investigate this question students among the various groups were compared on the performance levels that were dichotomized from a 5 level ranking scheme to two categories: Below the Standard and At or Above the Standard. The following figures indicate the relative performance among the 2 comparison groups.

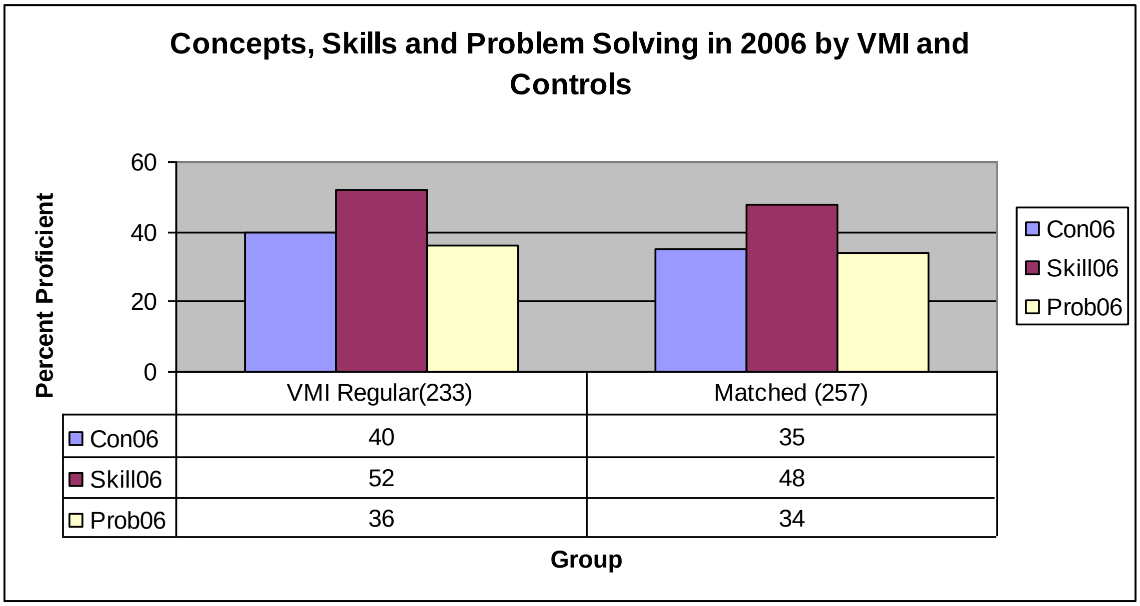
Title: The Vermont Mathematics Initiative:
Student Achievement from Grade 4 to Grade 10



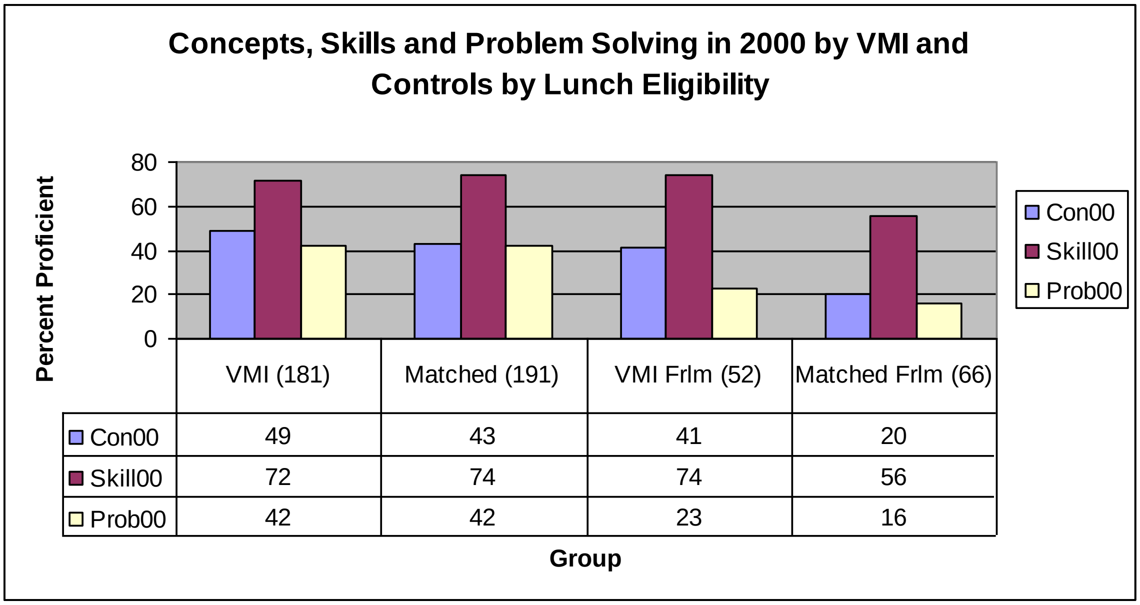
	VMI Regular(233)	Matched (257)
Con00	47	41
Skill00	73	69
Prob00	47	36



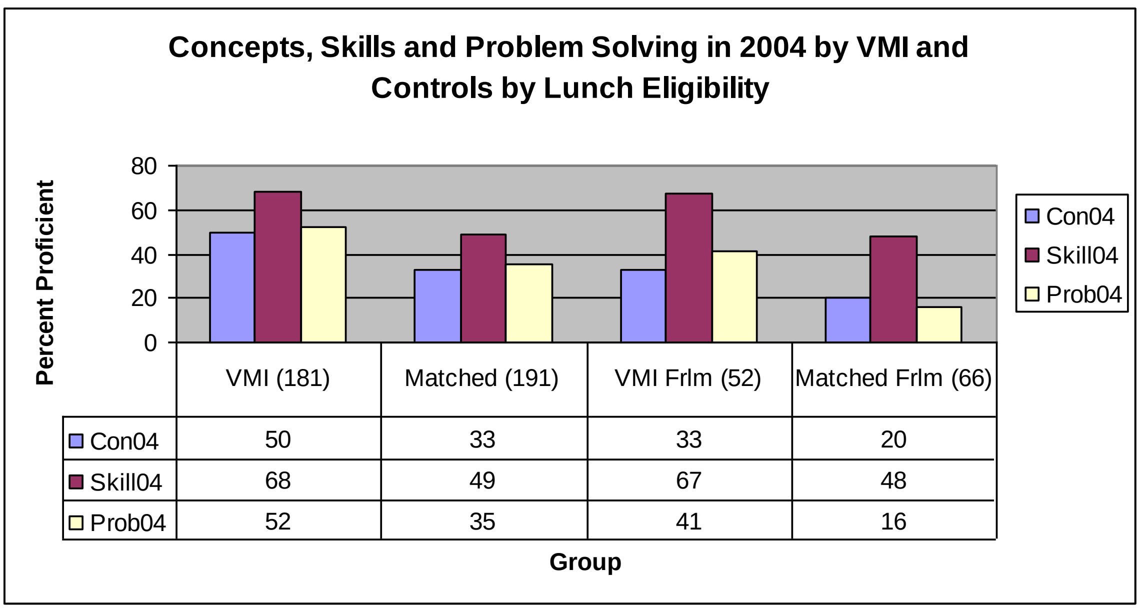
	VMI Regular(233)	Matched (257)
Con04	68	52
Skill04	55	44
Prob04	50	36



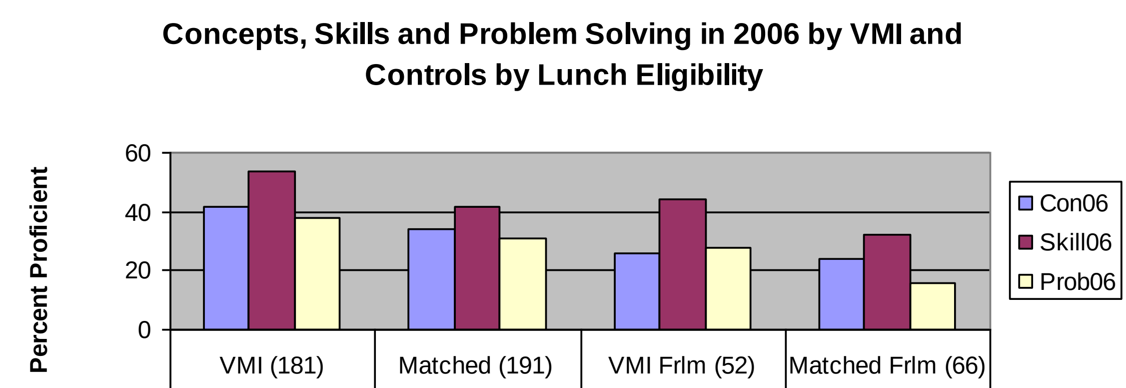
Patterns of performance between the groups are similar from year to year. Percentages from year to year, however are not comparable due to differences in scoring of the constructed response items. Patterns of relationships among the performance categories are probably a reflection of the particular items within years and the scoring protocols during each of the years. In the year 2000, the cohort differences were minimal and not statistically significant. At the 2004 data point, however the VMI group was significantly different from the Control group on all three measures. Skills: $x^2=12.83$, df 1, $p<.001$; Concepts: $x^2=6.36$, df1, $p<.001$; Problem Solving: $x^2=9.60$, df 1, $p<.001$. Within the 2006 year comparisons the performance of both the intervention and control groups was similar, and the differences which favor the VMI group were not statistically significant. The following figures illustrate the performance across years for the VMI and Control Groups when disaggregated by free or reduced lunch eligibility.



	VMI (181)	Matched (191)	VMI Frlm (52)	Matched Frlm (66)
Con00	49	43	41	20
Skill00	72	74	74	56
Prob00	42	42	23	16



	VMI (181)	Matched (191)	VMI Frlm (52)	Matched Frlm (66)
Con04	50	33	33	20
Skill04	68	49	67	48
Prob04	52	35	41	16



	VMI (181)	Matched (191)	VMI Frlm (52)	Matched Frlm (66)
Con06	42	34	26	24
Skill06	54	42	44	32
Prob06	38	31	28	16

Group

As with the overall performance, patterns of performance among the groups formed by lunch eligiblity are similar from year to year. In general, the Frlm groups are significantly lower than their within-group peers. The 2000 matched Frlm group is significantly lower than the other three groups, a result that is similar to the pattern of the scale scores above. The 2006 results are among the most interesting because a pattern which began in 2004 showed that the VMI Frlm group was not significantly different in performance from the Matched non-eligible group in 2006 at grade 10 (a 3 percent difference).

- Is there a relationship between self reported measures of opportunity to learn mathematics at grades 4 and 8 between groups of students who were taught by teachers who attended and completed the VMI training program in either 1999 or 2000 and a random sample of those who were not taught by VMI teachers?

In order to investigate these relationships several measures of opportunities to learn mathematics included as questions on the New Standards Reference Examination annual testing of all students at grades 4, 8 and 10. Appendix ( ) lists these questions and codes. Of particular interest to the study were the following items:

- **About how often have you done each of these activities in Math this year?**

  -Worked on problems that can be solved in more than one way
  -Worked on problems that used a representation (such as a chart or graph)
  -Wrote about how you solved a problem
  -Made an oral presentation about your math work
  -Worked on math in small groups
  -Used your own ideas or classmates' suggestions to change your work and make it better (conferencing)
  -Reviewed your portfolio work with your teacher, one-on-one
  -Received written comments back from your teacher about your work (that used the math scoring criteria)
  -Rewrote or revised your work to make it better
  -Used a scoring guide or rubric to assess your own work
  -Worked on math with a parent or adult other than your teacher
  -Used manipulative (visual aides such as tiles, cubes, or models)

- **How many pieces of work do you have in your math portfolio so far this year?**
- **I have kept a math portfolio for 0, 1, 2, 3, 4, 5 or more years.**

The results from cross tabulation of these items indicated that students in general tended to report that they did each of these activities about once a week in grade 4. There was no significant difference between the VMI and Control groups with respect to any of the mathematics opportunity to learn items measured at grade 4. Unfortunately the same items were not measured again 4 years later. However, one index of possible difference might have been the number of portfolio pieces students reported in their portfolios in 2000 at grade 4 and the number of years they had been keeping portfolios by grade 8.

With respect to the number of portfolio pieces students kept in portfolios at grade 4, the median number was four pieces for both the VMI and Control groups. About 80 percent of both groups reported keeping portfolios. In 2004, grade 8, 93 percent of the total sample, both VMI and Controls reported keeping math portfolios for at least one year. The median number of years keeping portfolios was 4 years, but there was no difference in the proportions above and below the median between the VMI and Control groups

Discussion

Scale score comparisons.

The results of analysis of groups of students formed by their exposure to teachers who were intensively trained by the VMI program and those who were randomly selected from a population that was similar in socio-economic background indicated that the VMI group of students significantly out performed the control group on standard measures of mathematics achievement taken over an 8 year period. The groups were not statistically different at the first year of intervention, 2000, but by 2004 in grade 8 they were out performing their peers on the scale scores. This pattern of performance difference on the scale scores persisted through grade 10. To some extent the performance difference at grade 10 was a reflection of the difference in drop-out incidence between the VMI and Control groups which favored the VMI group by 20 percent. Both the VMI and Control groups significantly improved over the 8 year period on scale scores, consistent with the overall state gain in scores. To some extent the gains may have been mediated by a ceiling effect whereas the state median scored at the seventy-second national percentile.

Performance Level comparisons.

Expected differences in performance on measures of skills, concepts and problem solving not present at grade 4, like the scale score differences, emerged at grade 8 but did not likewise persist to grade 10. Lack of standardization in scoring and scaling of the three measures make longitudinal comparisons inappropriate. However, the ‘within-year’ independent group comparisons at grade 10 remain unexplained.

Free Lunch Eligibility comparisons.

An effect that appears to narrow the gap between low-income and other students who are the beneficiaries of VMI instruction is consistent with previous study of the VMI 1999 to 2005 cohort of students. The 1999 study used schools as a unit of sampling and defined VMI intervention schools according to whether one or more teachers were attending the VMI during the first 4 years of the study. VMI schools were then matched to other schools that shared similar free lunch eligibility distributions. While the 1999 study was not able to achieve the same level of control of the matching across years and teachers that the present study achieved, the results were similar. The gap between free lunch eligible students and others in the VMI schools reduced over time to an insignificant difference by grade 10.

Opportunity to Learn.

When the New Standards Reference Examination was adapted and then adopted by the State Education Department from the NSRE developed at Pittsburg a number of questions were added to attempt to measure, by student self report, certain opportunities to learn including opportunities to learn mathematics. The questions proved to have acceptable reliability and significantly discriminated groups of students formed by free lunch eligibility and others with respect to keeping mathematics portfolios and parents attendance at conferences. Distributions of student responses to the questions included in this study were expected and, for ranked items essentially normal. However, these questions tended not to discriminate between groups students formed by the VMI intervention and others. One serious limitation of this analysis was that while questions were tested in several subsequent years, as it happened the same questions were not asked in 2000, 2004 and 2006. Nevertheless, it appears that with respect to the portfolio process, where we might have expected some difference in student experience in grade 4 or 8, there was apparently no difference in student experience between the groups. Differences in student experience that may be due to VMI teacher instruction will have to wait for a qualitative inquiry.

Summary.

The purpose of the paper presentation was to present findings from Year 8 of an eight year study of the process and value added effects of a statewide, content rich, mathematics professional development program known as the Vermont Mathematics Initiative (VMI). Specifically, the purpose of this investigation was to discover whether or to what extent the outcome measures used by the State of Vermont to report student achievement in mathematics discriminated between groups of students who were taught by teachers who attended the VMI coursework. The results of the analysis of student test scores suggests that students who are taught by VMI trained teachers perform better on standard measures of mathematics performance; and, these students are less likely to drop out of school than other students.

Title: The Vermont Mathematics Initiative:
Student Achievement from Grade 4 to Grade 10

Educational or Scientific Importance of the Study

The National Research Council, (2004) US Department of Education (2004) and researchers at RAND (2004, 2006) have cited the need for better evaluations of teacher professional development programs. The combination of mixed method approaches with multiple strategies such as longitudinal studies of students acting as their own controls over time is the present attempt to respond to current concerns. In addition, intervention students in combination with Control groups at multiple intervals are a further attempt to provide additional confidence in results.

From the perspective of reform of how mathematics should be taught, the content of the VMI curriculum may represent the kind of prioritization of content that provides a focus for teachers that enables them to get beyond what Hiebert, et al (2005) term the "feature by feature" approach to improving teaching. The intensive focus on how teachers and students learn the fundamental (as distinguished from "basic") concepts of mathematics also may enable the integration of the "what" of mathematics teaching with the "how" of implementation. The extent to which it can be shown that such a focus is related to higher levels of achievement will inform a more comprehensive solution to the problems associated with reform.

References

Galley, M. (2004). Teachers get heavy doses of advanced mathematics. Education Week. April 7.

Hiebert, J et al (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 video study. *Educational Evaluation and Policy Analysis.* 27(2), 111-132.

Le, V., Stecher, B., Lockwood, J.R., Hamilton, L., Robyn, A, Williams, G.R., Kerr, K.A., Martinez, J.F. & Klein, S. (2006). Improving mathematics and science education: A longitudinal investigation of the relationship between reform-oriented instruction and student achievement. Santa Monica, CA: RAND Corporation.

Meyers, H. & Harris, D. (2005). Evaluating the Vermont Mathematics Initiative in a value added context. Paper presented at the annual meeting of the American Educational Research Association, Montreal, P.Q. April, 2005.

Meyers, H. & Brewer, R. "The Vermont Educational Assessment: Phase 1, Design and Pilot." Paper presented at the Annual Meeting of American Educational Research Association, San Francisco, CA: 1989.

McCaffery, D.F., Lockwood, J.R., Koretz, D. M. & Hamilton, L.S. (2004) Evaluating value-added models for teacher accountability. Santa Monica, CA: RAND Corporation.

National Research Council. (2004) On evaluating curricular effectiveness: Judging the quality of k-12 mathematics evaluations. Washington, DC: The National Academies Press.

Raudenbush, S. (2005). Learning from attempts to improve schooling: The contribution of methodological diversity. *Educational Researcher*. 34(5), .25-31.

RMC Research Corporation. (2004). State mathematics and science partnerships: Year 1 implementation. Boulder, CO: RMC Research Corporation.

Stecher, B. M. & Mitchell, K.J. (1995). Vermont teachers' understanding of mathematical problem solving and "good" math problems. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, April 18-22, 1995.

Stuart, E. (2007). Estimating causal effects using school-level data sets. *Educational Researcher*. 36(4), 187-198.

Title: The Vermont Mathematics Initiative:
Student Achievement from Grade 4 to Grade 10

Appendix: NSRE OTL Questions for Vermont in the year 2000

About how often have you done each of these activities in Math this year? -Worked on problems that can be solved in more than one way -Worked on problems that used a representation (such as a chart or graph) -Wrote about how you solved a problem -Made an oral presentation about your math work -Worked on math in small groups -Used your own ideas or classmates' suggestions to change your work and make it better (conferencing) -Reviewed your portfolio work with your teacher, one-on-one -Received written comments back from your teacher about your work (that used the math scoring criteria) -Rewrote or revised your work to make it better -Used a scoring guide or rubric to assess your own work -Worked on math with a parent or adult other than your teacher -Used manipulative (visual aides such as tiles, cubes, or models)	**Alpha (A = More than Once a Week** **B = About Once a Week** **C = Once or Twice a Month** **D = 1-4 times This year** **E = Never) ,** **Blank or +** **(multiple marks)**
How many pieces of work do you have in your math portfolio so far this year?	**Numeric 00 = 0** **01 = 1** **02 = 2** **03 = 3** **04 = 4** **05 = 5** **06 = 6** **07 = 7** **08 = 8** **09 = 9** **10 = 10** **11 = 11 or more**
I have kept a math portfolio for 0, 1, 2, 3, 4, 5 or more years	**Numeric (0-5) or Blank**