July 22, 2003

Expansion Schools Report #3: Assessing Instructional Practices in eMINTS Classrooms

The eMINTS evaluation has highlighted the role of inquiry-based instructional practices in supporting high levels of student achievement. As the eMINTS project has developed, "instructional practice" has been assessed in three ways: through direct observation using the eMINTS Lesson Typology, through coding each lesson along Bloom's Taxonomy of Educational Objectives, and as part of a general assessment of the classroom environment using the "Hallmarks of an Effective eMINTS Classroom" rubric. This paper assesses the validity of each of these measures by analyzing their interrelationships. Data for this analysis is taken from a set of classroom observations conducted with teachers participating in the eMINTS expansion project. Results show significant correspondence between each measure at both the lesson and classroom levels.

During its first two years, the eMINTS program has proven to be successful in transforming elementary school classrooms into stimulating, interactive, high technology environments for learning. eMINTS has provided technological resources for classrooms, improved teacher's effectiveness in the classroom through significant professional development activities, and improved Missouri Assessment Program (MAP) test scores of eMINTS students. In analyses of 2001 and 2002 MAP scores, eMINTS students in classrooms characterized by the implementation of inquiry-based instructional practices scored higher than eMINTS students in classrooms characterized by more traditional, Teacher-centered instruction.[1]

In each of these analyses, teacher instructional practices were assessed in a different way. In the 2001 analysis, members of the evaluation team assessed instructional practices through the direct observation of classroom lessons. In the 2002 analysis, in addition to classroom observation, classroom instructional practices were assessed through a general rubric developed by the program professional development staff. The focus of this report is the consistency, face validity and inter-rater reliability of these assessments based on classroom visits to a set of 27 schools participating in the eMINTS expansion project.

Methods

The eMINTS expansion project was an experiment into the ways that eMINTS resources could be leveraged to support whole school reform. During the third year of the program,

[1] See *Analysis of 2001 MAP Results for eMINTS Students* and *Analysis of 2002 Results for eMINTS Students* on the eMINTS website http://emints.more.net/evaluation.

27 "expansion schools" were selected from the first and second cohorts of the project in order to explore the impact of eMINTS within individual schools. The 25 districts represented by the eMINTS expansion project received two additional eMINTS classrooms and professional development for two additional eMINTS teachers in the third and/or fourth grade.

The evaluation of the eMINTS expansion project is scheduled to cover two years. In order to collect baseline data, the eMINTS evaluation team members visited all expansion schools, observing the classrooms of 99 teachers in the winter of 2002. Individual lessons and instructional activities were coded on two scales, the eMINTS Lesson Typology[2] and Bloom's Taxonomy of Educational Objectives[3].

After the classroom observations, each teacher was interviewed regarding his or her activities as an eMINTS teacher. Using this information and information obtained from the classroom observation, team members assessed the teacher's overall instructional methods using a rubric designed by the eMINTS professional development team titled "Hallmarks of an Effective eMINTS Classroom." This instrument is a summary of instructional practices.

Content of Report

The analysis below is conducted on two levels. The first analysis focuses on individual lessons by comparing a score derived from Bloom's Taxonomy of Educational Objectives with the classification based on the lesson typology. The second analysis compares the total score on the "Hallmarks of Effective eMINTS Classroom" (the "eMINTS Hallmarks" score) to a summary classification of instructional practices based on the eMINTS Lesson Typology. Both analyses show a high level of consistency of across the different assessments of instructional practice.

The last section of the analysis assesses the inter-rater reliability of the eMINTS Hallmarks score. This analysis uses classification of classroom scores from four sources: the evaluation team, the eMINTS Cluster Instructional Specialists, the eMINTS building principal and the individual teacher. The inter-rater reliability is relatively low, highlighting the different interpretations of the rubric by different groups of raters.

Assessment of eMINTS Lesson Activities: eMINTS Lesson Typology and Bloom's Taxonomy

The first analysis considers lesson activities using two measures, the classification of the lesson in terms of the eMINTS Lesson Typology and a score derived from the lesson's Bloom's Taxonomy rating. Members of the evaluation team made both of these determinations during their two-hour observation of classroom instruction.

---

[2] See *A General Typology of eMINTS Lessons* on the eMINTS website http://emints.more.net/evaluation.

[3] Bloom, B. S. (Ed), (1956), *Taxonomy of educational objectives; The classification of educational goals*. New York: Longmans Green.

Table 1
Distribution of eMINTS Lesson Typology Categories, Observed Lessons

	Number of Lessons	Percentage
Teacher-centered	100	42.4
Hybrid	60	25.4
Student-centered, Facilitated	76	32.2
Total	236	100.0

eMINTS Lesson Typology

The first measure of lesson activities was the eMINTS Lesson Typology. This measure of instructional practice was originally developed in 2001 to classify the key characteristics of inquiry-based lessons. These lessons, referred to as "Student-centered, Facilitated" lessons, are structured to give students the opportunity to research and investigate a lesson's topic. There are two other main categories in this typology: "Teacher-centered lessons", which correspond to traditional, teacher led instruction, and "Hybrid lessons", which is a mixed type of lesson.[4] The distribution of lessons by eMINTS Lesson Typology category is presented in Table 1.

Evaluation team members observed classroom instruction as part of the school visits conducted in the winter of 2002. Evaluators coded each major lesson activity using each scale. Evaluators observed a total of 236 activities in the 99 classrooms they visited. In each classroom, between one and five distinct lesson activities were observed, with an average number of 2.4 observed activities per teacher.

The results in Table 1 show that slightly less than one-third of observed lessons were classified as "Student-centered, Facilitated."

Bloom's Taxonomy of Educational Objectives

The second measure of lesson activities was Bloom's Taxonomy of Educational Objectives. This taxonomy is a conventional measure of educational objectives based on a hierarchy of cognitive tasks. This continuum ranges from the recall and recitation of simple facts to the interpretation and discrimination of complex ideas. The evaluation team members used the six levels of Bloom's classification system to rate each observed lesson's level of abstraction. The distinctions in the character of knowledge expected by teachers are embodied in the six categories of the taxonomy. These categories are labeled in Table 2.

[4] See the eMINTS report, *A General Typology of eMINTS Lessons*, for a full description of the typology.

Table 2
Distribution of Bloom's Taxonomy Scores, Observed Lessons

Score	Category	Number of Lessons	Percentage
1	Knowledge	85	36.0
2	Comprehension	49	20.8
3	Application	56	23.7
4	Analysis	21	8.9
5	Synthesis	18	7.6
6	Evaluation	7	3.0
	Total	236	100.0

The distribution of lesson codes in Table 1 is clearly skewed towards the lower end of Bloom's Taxonomy. The category with the largest number of entries was the lowest category, "knowledge". Examples of this type of lesson included classroom drill and worksheet completion.

The Relationship Between Bloom's Taxonomy and Inquiry-Based Lesson Activities

The instructional leaders of the eMINTS program maintain that the inquiry-based approach to instruction will produce lessons that fall within the upper half of Bloom's Taxonomy. In other words, as lesson tasks become more integrative and complex, they engage in more complex cognitive tasks and higher-order learning activities.

The distribution of lessons by the eMINTS Typology categories is presented in Table 3. This table collapses the 6 categories of the Bloom's Typology into two general categories. The first category, from "Knowledge" to "Application", represents the lower-order types of learning activities. The second category, from "Analysis" to "Evaluation", represents the higher-order learning activities encouraged in the eMINTS instructional program. Table 3 clearly shows 46.1 percent of lessons classified as "Student-centered, Facilitated" (i.e., inquiry-based lessons) were also classified in the upper half of Bloom's Taxonomy. In comparison, fewer than ten percent of "Teacher-centered" and "Hybrid" lessons applied these higher-order cognitive activities.

In light of this relationship, one would also expect to see differences in the observed Bloom's Taxonomy Score by categories of the eMINTS Lesson Typology. In particular, one would expect to see higher average scores on the Bloom's Taxonomy scale as lessons become "more" inquiry-based. This is seen in Table 4. Table 4 presents the mean values of the Bloom's Taxonomy across different categories of the eMINTS Lesson Typology. A lesson's score on the Bloom's Taxonomy is equal to its level, thus a lesson that falls into the "Knowledge" category would receive a score of "1", while a lesson that falls into the "Evaluation" category would receive a score of "6". The table shows significant differences in the Bloom's Taxonomy score across each category of the eMINTS Lesson Typology. Overall, the relationship between the Bloom's Taxonomy score and the lesson typology category is moderately strong. The variation across the categories of the lesson typology accounts for 29 percent of the variation in the Bloom's Taxonomy score.

Table 3
Distribution of Levels of Bloom's Taxonomy by eMINTS Lesson Typology

	eMINTS Lesson Typology				
Level of Bloom's Taxonomy	Teacher-centered	Hybrid	Student-centered, Facilitated	Total	Number of Lessons
Knowledge to Application	93.0	93.3	53.9	80.5	190
Analysis to Evaluation	7.0	6.7	46.1	19.5	46
Total	100.0	100.0	100.0	100.0	236
P-Value	<0.0001				

Table 4
Mean Values of Bloom's Taxonomy by eMINTS Lesson Typology, Observed Lessons

Lesson Typology	Number of Lessons	Mean	Standard Deviation
Teacher-centered	100	1.70	1.18
Hybrid	60	2.27	1.04
Student-centered, Facilitated	76	3.43	1.31
Total	236	2.40	1.40
P-value	<0.0001		
R-Squared	0.29		

Differences in Means (**Bold** Differences p <= .05)

	Teacher-centered	Hybrid	Student-centered, Facilitated
Teacher-centered	0.00		
Hybrid	**0.57**	0.00	
Student-centered, Facilitated	**1.73**	**1.17**	0.00

The mean score of the Bloom's Taxonomy is significantly higher in each category of the lesson typology. The magnitude of these differences is striking. For example, the average score for Student-centered, Facilitated lessons is 1.73 points higher that the score for Teacher-centered lessons. This difference is larger than the overall standard deviation in the Bloom’s Taxonomy score.

Summary

The analysis of the codes given to observed lessons shows two important relationships. First, the observations of the eMINTS professional development staff are supported, namely that inquiry-based lessons are more likely than other types of lessons to employ higher-level thinking skills, as measured by Bloom's Taxonomy.

The second point is the existence of a moderately strong linear relationship between the Bloom's Taxonomy score and the eMINTS Lesson Typology category. This relationship suggests that, at the level of individual lessons, that both classification systems are measuring similar lesson characteristics. Given that Bloom's Taxonomy is an accepted classification system for instructional activities, the relationship between it and the eMINTS Lesson Typology supports the face validity of the lesson typology.

Overall Assessment of eMINTS Classrooms: The eMINTS Hallmarks and the Aggregate eMINTS Lesson Typology

The previous analysis considered the classification of observed lessons. Other information collected by the evaluation team, along with data collected by eMINTS Cluster Instructional Specialists, building principals, and teachers, refers to entire classrooms. This document, the "Hallmarks of an Effective eMINTS Classroom" is a general rubric assessing the application of the eMINTS instructional model in individual classrooms. This section describes the relationship between a summary measure derived from the eMINTS Lesson Typology and the eMINTS Hallmarks rubric. First, the method used to aggregate the eMINTS Lesson Typology to the classroom level is described. Second, the characteristics of the "Hallmarks of an Effective eMINTS Classroom" are discussed. Finally, the relationship between these two classroom-level variables is considered.

Aggregating the eMINTS Lesson Typology

As mentioned above, a total of 236 distinct lesson activities were observed during 99 classroom visits. These lessons included a broad range of activities and subjects, from the completion of worksheets to long-term WebQuests. Each observed activity was coded according to the eMINTS Lesson Typology. To analyze the relationship between the lesson typology and other classroom-level measures, lesson activities were aggregated by teacher. Classrooms were classified by their most frequent lesson category. The summary classroom-level classifications are presented in Table 5.

Table 5
Distribution of eMINTS Lesson Typology Categories, eMINTS Classrooms

	Number of Classrooms	Percent
Teacher-Centered	32	32.3
Hybrid	25	25.2
Student-centered, Facilitated	42	42.4
Total	99	100.0

Hallmarks of an Effective eMINTS Classroom

The second measure is a measure of eMINTS classrooms developed by the eMINTS professional development staff. The "Hallmarks of an Effective eMINTS Classroom" is a rubric for teachers and staff to use as a guide for the development of a student-centered classroom environment. Cluster Information Specialists, building principals, and teachers use this rubric to assess the overall application of inquiry-based and constructivist teaching practices in eight domains, ranging from the use of facilitated learning techniques to the application of constructivist principles in curriculum planning and assessment.[5] During their school visits, evaluation team members also rated classrooms using this rubric.

Scoring Classroom Hallmarks Ratings

Using the eMINTS Hallmarks, evaluation team members rated the classrooms after their classroom observations and interview with each teacher. The scores on each of the 8 domains of the rubric ranged from 0 (for a teacher in the *Emerging* level) to 4 (for a teacher at the *Advanced* level). This created an overall score with a possible range of 0 to 32 points. The total score was then divided into four levels, identifying *Emerging, Experimental, Transitional,* and *Proficient* eMINTS classrooms. The distribution of individual levels is presented in Table 6.

The Relationship Between the eMINTS Lesson Typology and eMINTS Hallmarks

Tables 7 and 8 show the relationship between the eMINTS Lesson Typology and the eMINTS Hallmarks at the classroom level. Table 7 shows the relationship in terms of the summary categories of the Hallmarks scale. Table 8 analyzes the overall Hallmarks score by categories of the Lesson Typology.

[5] A discussion of this evaluation instrument can be found in the appendices of the report, *Analysis of 2002 MAP Results for eMINTS Students*, on the eMINTS web site: http://www.emints.more.net/evaluation

Table 6
Distribution of eMINTS Hallmarks Categories, eMINTS Classrooms

Hallmarks Category	Number of Classrooms	Percent
Emerging (0 to 8)	50	50.5
Experimental (9 to 16)	19	19.2
Transitional (17 to 24)	24	24.3
Proficient (25 to 32)	6	6.1
Total	99	100.0

Table 7
Distribution of eMINTS Hallmarks Category by eMINTS Classroom Typology, eMINTS Classrooms

	Teacher-centered	Hybrid	Student-centered, Facilitated	Total
Emerging	93.8	66.7	9.3	50.5
Experimental	3.1	25.0	27.9	19.2
Transitional	3.1	8.3	46.5	24.3
Proficient	0.0	0.0	16.3	6.1
Total	100.0	100.0	100.0	100.0
P-Value	<0.0001			

The results in Table 7 show that classrooms where teachers' most frequent lesson type was classified as "Student-centered, Facilitated" were more likely than other classrooms to be classified as *Transitional* and *Proficient* on the Hallmarks scale. Overall, slightly less than two-thirds of Student-Centered classrooms were classified as *Transitional* or *Proficient*, compared to less than one-tenth of other classrooms. In addition, all of the *Proficient* classrooms were classified as Student-centered, Facilitated.

Table 8 shows mean differences in the total score calculated from the Hallmarks rubric. Here, differences between categories of the lesson typology account for 61 percent of the variance in the total Hallmarks score. These results also show that classrooms classified as "Student-centered, Facilitated" scored ten points higher than "Hybrid" classrooms and nearly fifteen points higher than "Teacher-centered" classrooms.

Table 8
Mean Values of the eMINTS Hallmarks Rubric by eMINTS Lesson Typology, eMINTS Classrooms

	Number of Teachers	Mean	Standard Deviation
Teacher-centered	32	2.98	4.08
Hybrid	24	7.62	4.90
Student-centered, Facilitated	43	17.94	6.34
Total	99	10.58	8.54
P-Value	<0.0001		
R-Square	0.61		

Differences in Means (**Bold** Differences p <= .05)

	Teacher-centered	Hybrid	Student-centered, Facilitated
Teacher-centered	0.00		
Hybrid	**4.64**	0.00	
Student-centered, Facilitated	**14.96**	**10.32**	0.00

Summary

This brief analysis of classroom-level scores and classifications of two classroom-level variables supports the face validity of each. These results suggest that both the aggregated Lesson Typology and the eMINTS Hallmarks rubric measure the same thing, namely, the implementation of inquiry-based instructional practices.

The final section of this paper assesses the inter-rater reliability of the overall Hallmarks score, comparing the evaluators' judgment of classroom effectiveness with that of the eMINTS Cluster Instructional Specialists, the building principals, and the eMINTS teachers participating in the expansion project.

Inter-rater Reliability of eMINTS Hallmarks Ratings

The Hallmarks rubric was originally designed by the Cluster Instructional Specialists not only for their use in evaluating eMINTS classrooms but also to provide a reference guide for continual improvement of eMINTS classrooms for Cluster Instructional Specialists, teachers, and principals. The completion of the eMINTS Hallmarks by all four groups of raters, members of the evaluation team, Cluster Instructional Specialists, building principals, and eMINTS teachers, allows for the assessment of the overall inter-rater reliability of the rubric. This section examines the agreement between these groups of raters.

Table 9
Frequency of eMINTS Hallmarks Rubric Categories by Rater

	Evaluation Team	Cluster Instructional Specialist	Principal	Teacher
Emerging	50.5	16.7	5.2	0.0
Experimental	19.2	45.5	44.8	42.6
Transitional	24.2	30.3	36.2	46.8
Proficient	6.1	7.6	13.8	10.6
Total	100.0	100.0	100.0	100.0
Number of Observations	99	66	58	47

Table 9 presents the overall distribution of Hallmark categories from each group of raters. The number of available observations varies markedly by type of rater. Members of the evaluation team were able to complete Hallmarks rubrics on all 99 expansion teachers, while fewer than half of the expansion teachers submitted self-ratings. These differences in the number of observations bias the assessment of inter-rater reliability, since it is not possible to definitively state the agreement between different raters. Nevertheless, it is useful to review the extent of agreement based on the available data.

Table 10 presents weighted Kappa statistics for each pair of raters. Overall, the categorical agreement between raters is low: between 20 percent between members of the evaluation team and teachers; 42 percent between principals and teachers. Overall, the lowest level of agreement is for members of the evaluation team, while the agreement between the Cluster Instructional Specialists, principals and teachers is slightly higher.

This lack of agreement between the evaluation team and the other raters may be due to the overall application of this rubric. The Hallmarks rubric is a summary measure of classroom environment and is designed to assess classrooms over an extended period of time. Members of the evaluation team made their assessment of the expansion classrooms based on a single visit and a limited number of classroom observations. On the other hand, the Cluster Instructional Specialists, building principals, and teachers were able to score classrooms based on a longer perspective of development. It may be the case that the ratings of the Cluster Instructional Specialists, building principals and eMINTS teachers reflect a more accurate perspective on the operations of eMINTS classrooms.

Table 10
Inter-Rater Reliability Assessment, Hallmarks Categories

	Evaluation Team	Cluster Instructional Specialist	Principal	Teacher
Evaluation Team				
Weighted Kappa	----			
P-Value				
Number of Observations	99			
Cluster Instructional Specialist				
Weighted Kappa	0.1969	----		
P-Value	0.0066			
Number of Observations	66	66		
Principal				
Weighted Kappa	0.2182	0.3865	----	
P-Value	0.0004	0.0002		
Number of Observations	58	43	58	
Teacher				
Weighted Kappa	0.1950	0.3090	0.4190	----
P-Value	0.0001	0.0016	0.0041	
Number of Observations	47	42	30	47

Table 11
Mean Values of Hallmark Rubric by Rater

	Number of Ratings	Mean	Standard Deviation
Evaluation Team	99	10.58	8.54
Cluster Information Specialist	66	15.18	6.61
Principal	58	17.59	6.00
Teacher	47	18.30	4.89
Total	270	14.55	7.70

Table 11 presents the mean value of the Hallmarks scores for different raters. It is clear that teachers give their classrooms the highest average scores, while members of the evaluation team gave the expansion classrooms the lowest. This difference probably reflects several things. Teachers may bias their ratings upwards to account for their learning and development within the eMINTS programs. On the other hand, the evaluation team members clearly do not rate the classrooms they visit with the same perspective as do the Cluster Instructional Specialists, building principals or teachers. A

difference in the interpretation of the Hallmark categories could also account for a difference in the scoring between raters.

Conclusions

The analysis discussed above is an overall assessment of approaches to measuring instructional practices in eMINTS classrooms. The results compare lessons, through a comparison between Bloom's Taxonomy and the eMINTS Lesson Typology, and classrooms, through a comparison between an aggregate measure derived from the eMINTS Lesson Typology and the "Hallmarks of an Effective eMINTS Classroom" rubric. Specifically, the results show consistency in measurement at both the lesson and classroom levels. However, the inter-rater reliability on the Hallmarks rubric between members of the evaluation team, Cluster Instructional Specialists, building principals and teachers is relatively low.

There is a measure of consistency between the Bloom's Taxonomy, the eMINTS Lesson Typology, and the "Hallmarks of an Effective eMINTS Classroom" rubrics when applied by the evaluation team. This suggests that these measures have some face validity in their assessment of instructional and classroom practices.

In previous analyses of MAP scores, high levels of student performance is related to the application of inquiry-based instructional practices. This is seen regardless of whether instructional practice is measured through direct observation (using the eMINTS Lesson Typology) or through a more general rubric of classroom operation (using the eMINTS Hallmarks rubric).

The current analysis shows a positive linear relationship between standard educational assessments (e.g., the Bloom's Taxonomy score) and the more interpretative assessment of lesson practices developed by the evaluation team (the eMINTS Lesson Typology). This relationship highlights one of the key assertions of the eMINTS instructional model, that the implementation of lessons that require higher-order thinking skills (i.e., lessons in the upper half of Bloom's Taxonomy) taught in a student-centered, constructivist framework (as measured by the eMINTS Lesson Typology) support high performance on the MAP test.

At the classroom level, the analysis suggests that the characteristics embodied in the Hallmarks rubric are positively related to the overall conduct of eMINTS lessons. Both the Hallmarks categories and Hallmarks scores are higher for classrooms characterized by Student-centered, Facilitated instruction. This suggests that both instruments measure the application of inquiry-based instruction in eMINTS classrooms. As seen in the two MAP reports, eMINTS students in classrooms characterized by inquiry-based instruction scored higher on the MAP tests than did other eMINTS students. The present analysis, as well as the two MAP reports, suggests that the eMINTS Lesson Typology and the eMINTS Hallmarks rubric are valid measures of the application of the overall eMINTS instructional model.

While the various measures of instructional practices show some consistency, there is no real inter-rater reliability for the Hallmarks rubric. The extent of agreement among Hallmarks categories is relatively low. There are several potential reasons for this, everything from differences in the context of the rating, to different interpretations of the rubric's content. Further refinements of the Hallmarks rubric may be necessary to achieve more acceptable levels of inter-rater reliability.

Despite the low inter-rater reliability of the Hallmarks rubric, the analysis presented here demonstrates the basic validity of the two main measures of instructional practice used by the eMINTS program. Consequently, these results suggest that the use of these measures, the eMINTS Lesson Typology and the Hallmarks of an Effective eMINTS Classroom, should continue.