DOCUMENT RESUME

ED 442 879 TM 031 305

AUTHOR Cerrillo, Tracy L.; Hansen, Mary A.; Parke, Carol S.; Lane, Suzanne; Scott, Kathryn

TITLE The Relationship between MSPAP and Science Classroom Instruction and Assessment Materials.

SPONS AGENCY Department of Education, Washington, DC.
PUB DATE 2000-04-00
NOTE 31p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 25-27, 2000). For the evaluation of Maryland teachers' social studies materials, see TM 031 304.
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Education; *Elementary School Teachers; *Instructional Materials; *Middle School Teachers; Middle Schools; Outcomes of Education; *Performance Based Assessment; *Science Instruction; Sciences; *Student Evaluation
IDENTIFIERS Maryland; *Maryland School Performance Assessment Program

ABSTRACT

As part of a research project being conducted to determine the impact that the Maryland Learning Outcomes and the Maryland State Performance Assessment Program (MSPAP) are having on classroom instruction and assessment practices, science classroom materials were collected from elementary and middle schools in Maryland during the 1997-1998 school year. These materials were examples of instruction and assessment activities used by teachers in their day-to-day teaching. The activities were analyzed with respect to the extent to which they are aligned with the Maryland Learning Outcomes in science and the extent to which they represent MSPAP tasks with regard to such features as processes assessed, format, response types required, and integration with other subject areas. In all, 301 science teachers from 62 schools sent in social studies activities, a response rate of 70%. A coding scheme was developed to reflect important features of these tasks. Instruction and assessment activities showed different patterns. In general, instruction activities were more similar to the MSPAP than assessment. The findings suggest that the instruction, assessment, and test preparation activities sent in by Maryland teachers reflect the Maryland Learning Outcomes from a moderate to high degree. However, instruction and assessment activities are less similar to the Maryland Learning Outcomes and MSPAP than the test preparation activities. Results from related studies of alignment in the subject areas of reading, mathematics, writing, social studies, and science are discussed for their similarities and differences in the degree of alignment. (Contains 16 tables.) (SLD)

The Relationship Between MSPAP and Science Classroom Instruction and Assessment Materials

by

Tracy L. Cerrillo, Mary A. Hansen, Carol S. Parke, Suzanne Lane, Kathryn Scott

University of Pittsburgh

Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, April 2000. Preparation of this paper was supported by a grant from the U. S. Department of Education, Assessment Development and Evaluation Grants Program (CDFA 84.279-A), for the Maryland Assessment System Project. Our deepest appreciation is extended to the teachers, principals, and students in Maryland for their invaluable time and effort spent on this project.

The Relationship Between MSPAP and Science Classroom Instruction and Assessment Materials

Educational reform policy has recently incorporated a strategy that involves the use of setting and developing national standards. The first set of standards appeared in 1989 from the National Council of Teachers of Mathematics (1989). Since then, standards have been developed in other subject areas as well. The National Science Education Standards (1996) were developed under the premise that all students deserve the opportunity to achieve scientific literacy. They outline the content and scientific processes that students should know in order to empower all students in scientific inquiry. These standards envision a change in emphasis in the way that science is presented and taught to students. The National Science Education Standards (1996) place more emphasis on communicating science explanations, applying the results of experiments, and public communication of student ideas, and less emphasis on getting an answer.

Implementing the National Science Education Standards (1996) will require changes in science educational practices in much of the nation's states and districts. The standards can be organized and presented in many different ways, which allows local, state, and national levels to judge which aspects of the standards will serve the needs of their particular students and/or communities. Many states have used these standards as the basis for developing specific learning goals for their own students. One such state is Maryland, which began its implementation of a statewide performance assessment in the early 1990's. The Maryland State Performance Assessment Program (MSPAP) is a performance assessment program that is designed to measure school performance for grades 3, 5, and 8 in the state of Maryland (Maryland State Board of Education, 1995).

The Maryland State Department of Education, in conjunction with content educators in science, determined the learning outcomes and indicators that MSPAP would assess for the science subject area. These learning outcomes deal with the understanding of content and abilities related to science and are assessed by the MSPAP test questions. MSPAP was intended to promote performance-based instruction and assessment practices based on the Maryland Learning Outcomes.

As part of a research project being conducted to determine the impact that the Maryland Learning Outcomes and MSPAP itself are having on classroom instruction and assessment practices, classroom materials were collected from elementary and middle schools in the state of Maryland. These classroom materials were examples of instruction and assessment activities used by teachers in their day-to-day teaching. The classroom activities were analyzed with respect to the level in which they assess the Maryland Learning Outcomes in the science subject area. This paper presents the findings of the

analyses. It should be noted that many of the results presented in this paper do not include grade level results. This is because the differences in the percentages across grades were negligible.

Method

Sample

Science teachers were asked to send in a sample of their science classroom activities used during the 1997-98 school year. Activities were collected in the fall and in the spring. Science teachers were asked to send in approximately 3 to 4 instruction activities, 3 to 4 assessment activities, and 1 sample of a scoring scheme used during each time period. In addition, in the spring, 3rd, 5th, and 8th grade science teachers were also asked to send a sample of a MSPAP test preparation activity used prior to the administration of MSPAP.

A data collection form was developed to obtain information from each teacher regarding the classroom from which the activities were selected. The data collection forms asked teachers to indicate the grade level, the nature of the students' ability levels for the science class (e.g., heterogeneous ability group, homogeneous ability group, exclusively special education, exclusively gifted and talented), and the nature of the content taught in the class (e.g., general science, life science, physical science, etc.). Each teacher completed a form and returned it along with their activities for each collection period.

Overall, 74 schools with 427 teachers were asked to participate in this aspect of the study. Some or all of the teachers from 62 of the schools participated, resulting in a school participation rate for classroom activities of 84%. Within these 62 schools, 301 science teachers sent in all or a subset of the activities requested (70%). Most of the elementary classes (85%) and middle school classes (65%) were 'general science'. For most of the remaining elementary school classes (14%) and middle school classes (13%) the teacher did not indicate the type of class on the label provided. There was an additional set of middle school classes that were classified as 'physical science' (8%), 'life science' (8%), 'other' (6%).[1]

Teachers also indicated the heterogeneity of the students in the science class from which their sample of classroom activities were selected. As indicated in Table 1, 74% of the elementary classes, compared to only 48% of the middle school classes, were classified by the teacher as heterogeneous. A larger percentage of middle school classes (22%) than elementary classes (11%) were classified as homogenous, either on-grade, above-grade, or below-grade level.

[1] Other represents earth science, health science, chemistry, biology, or 'dynamic earth' classes.

Table 1.
Heterogeneity of Science Class from which the Classroom Materials were Selected

	All grades n=357*	Elementary n=227	Middle n=130
Heterogeneous	65%	74%	48%
Homogeneous	15%	11%	22%
On-grade	9%	5%	17%
Above-grade	3%	2%	4%
Below-grade	3%	4%	1%
Exclusively Special Education	7%	5%	10%
Exclusively Gifted/Talented	3%	1%	5%
Not indicated	11%	9%	15%

* Note: This number is larger than the 301 teachers who sent in classroom activities because 56 teachers had a change in the type of science class taught from fall to spring

Description of Classroom Activities

Teachers were provided with labels to attach to each activity indicating the type of activity (e.g., instruction, assessment, test preparation, scoring scheme). Table 3 shows the number and percentage of activities for each type. Across all grades (2, 3, 4, 5, 7, and 8), there was a total of 1,666 instruction activities, 1,174 assessment activities, and 810 scoring schemes. For grades 3, 5, and 8 there was a total of 125 MSPAP test preparation activities. The table also includes a category called 'Not Coded'. These were activities that were not coded for one of two reasons. One reason for not coding an activity was because it pertained strictly to another content area, such as social studies or mathematics. Another reason an activity was not coded was because it consisted only of teacher notes or general lesson plans, and it was difficult to discern what the students were required to do.

The last three columns of Table 3 indicate the number and percentage of teachers providing each type of activity, as well as, the average number of each type of activity per teacher. As an example, about 42% (1,666 out of 3,973) of the activities received were labeled as instruction, and 97% the teachers (293 out of 301) sent in at least one instruction activity.

Table 3.
Type of Classroom Activity

	Activities n= 3973		Teachers n=301		Mean Number of Activities Per Teacher
	Number	Percentage	Number	Percentage	
Instruction	1,666	42%	293	97%	5.69
Assessment	1,174	30%	266	88%	4.41
MSPAP Test Preparation	125	3%	66	40%*	1.89
Scoring Schemes	810	20%	253	84%	3.20
Not Coded	198	5%	99	33%	2.0

*Note: This percentage is based only on the 163 on-grade teachers (3, 5, 8) and not on the full sample of 301 teachers.

Sources of Classroom Activities

Science teachers indicated the source of each activity. Table 4 indicates the sources of the instruction, assessment, MSPAP test preparation activities, and scoring schemes across all grades. The results across grades were quite similar. Nearly one-third (32%) of the instruction activities were from textbook/commercial resources, and nearly another third (32%) were teacher-developed. However approximately half of the assessment activities were teacher developed (49%), and only 20% were from textbook/commercial resources. When examining the MSPAP test preparation activities, the sources were somewhat different than for the instruction or assessment activities. The percentage of teacher-developed activities was similar to instruction activities (29%). However, there was a larger percentage of test preparation activities that were county-developed (31%), MSPAP Release Tasks (4%) and other state-level materials (5%), as compared to instruction activities.

Table 4.
Sources of Classroom Activities

	Instruction n=1666	Assessment n=1174	MSPAP Test Prep. (3^{rd}, 5^{th}, 8^{th}) N=125	Scoring Scheme n=810
Teacher/Other Teacher/ School Developed	32%	49%	29%	44%
Textbook/Commercial Resources	32%	20%	16%	7%
County/Another County Developed	17%	13%	31%	27%
Teacher and Textbook/ Teacher and County Developed	9%	8%	6%	4%
MSPAP Release Tasks	<1%	<1%	4%	<1%
MD Consortium/ Exemplars	<1%	<1%	5%	1%
Other	2%	<1%	2%	2%
Cannot Be Determined	8%	8%	7%	15%

Procedure

The classroom instruction, assessment, and MSPAP test preparation activities were analyzed for each science teacher using a coding scheme designed to provide information about the format of the activities, the extent to which the activities reflect the Maryland Learning Outcomes for science, the overall similarity to MSPAP-like tasks, and a variety of other features (e.g., response type required of students, integration with other subject areas, etc.). The Maryland Learning Outcomes and the format and content of MSPAP served as the basis for the coding schemes that were developed for the analysis of the classroom activities.

A total of six raters coded the classroom activities. A formal training session was conducted to familiarize the raters with the coding scheme using a sample set of pre-coded activities. Then, the raters coded another set of sample activities independently and their codes were compared and discussed by the group. After the formal training was complete, pairs of raters individually coded a set of classroom activities from a school (elementary or middle) for a certain collection period (fall or spring). The pair of raters met to discuss their discrepancies and reached a consensus on the codes for each activity. This was done to ensure that all raters shared a common understanding of the coding scheme.

After it was determined that the raters reached a shared understanding of the coding scheme and were proficient in applying it to a variety of classroom activities, each rater individually coded sets of classroom activities. Approximately 20% of the sets of classroom activities (an elementary or middle school teacher's activities from either fall or spring) were coded individually by two raters. The overall adjusted rate of agreement between the raters was then calculated[2]. The adjusted rate of agreement was found to be 85% for the instruction, assessment, and test preparation activities.

Results

Maryland Process Learning Outcomes

The classroom instruction, assessment, and MSPAP test preparation activities were coded in terms of whether they focused on the Process Learning Outcomes as defined by the Maryland Learning Outcomes (MLO's) and MSPAP. The Process Learning Outcomes for science are Nature of Science, Habits of Mind, Attitudes, Processes of Science, and Applications of Science. It should be noted that Attitudes was not included in the coding scheme because the main purpose of the outcome is that students should develop a positive attitude toward science. It was determined that this outcome would be difficult to

[2] Percent agreement was considered to be too lenient of an index of rater agreement because for a number of the categories to be coded there were a range of options that could be selected. As an example, for the Process Learning Outcomes, one to nineteen process outcomes could be selected for an activity. However, the majority of the activities had between one to six process outcomes coded. A simple percent agreement based on each of the nineteen outcomes would have inflated the index for rater agreement. Thus, an adjusted percent agreement was used.

identify in classroom instruction and assessment materials, and thus was not coded. Nature of Science, Habits of Mind, and Process of Science were categorized into more specific processes that are indicators of these more general learning outcomes. The sub-categories were established by evaluating the indicators that were presented in the MLO's, and then by determining how each indicator should be manifested in the coding scheme. In some instances, the indicator presented in the MLO's could not be appropriately represented in a form that was codeable. Therefore, there is not a direct correspondence between the sub-categories under each process and the indicators as defined in the MLO's. Also, some of the indicators are associated with more than one learning outcome, so there is some degree of overlap. The Applications of Science category remained intact as defined in the MLO's, as opposed to being broken down by a number of indicators, because it was determined that the description provided in the MLO's allowed for accurate coding of the category. Below is a description of the Process Learning Outcomes and the indicators used in the coding scheme.

Five indicators were coded under the Nature of Science process outcome including 1) creating and/or modifying models, 2) determining and recognizing patterns, 3) making predictions and/or formulating hypotheses, 4) detecting bias and/or fairness, and 5) considering other's ideas or perspectives. Tasks were coded for model usage only if students were required to create a model or modify an existing model, and not if they simply had to use a model that was provided to them. Pattern recognition included analyzing simple graphs to determine patterns in data and/or examining experiences in order to demonstrate repetition or patterns in nature. The indicator, 'making predictions and/or formulating hypotheses' was selected when students were required to make predictions or hypotheses regarding experiments or natural phenomena. 'Detecting bias and/or unfairness' was chosen when students had to identify sources of bias in their investigations. In addition it was selected if students were asked questions related to experimental design issues, such as validity concerns of a particular experiment. It was also coded if students were asked to share results with other classmates in order to determine causes related to any differences in results that may have been obtained for different groups of students. 'Considering other's ideas or perspectives' was selected when students had to consider the perspective of other students or scientists. It was also coded when students were required to take a position on some scientific issue and defend it with a rationale. Peer responses to student work were also coded in this category because the purpose of peer response is to share ideas and have students critique and analyze others' work.

The four Habits of Mind indicators were: 1) ordering, classifying, and/ or sequencing objects, or drawing and labeling a diagram, 2) comparing and contrasting and/or determining similarities and differences, 3) designing and/or modifying experiments, and 4) posing scientific questions. It should be

noted that the first indicator is rather broad in terms of the types of skills that are included in the category. Ordering, classifying, and sequencing are skills that are often difficult to disentangle. Therefore, they were all included together. The other component in this category is 'draw and label a diagram'. This skill was included in this indicator because it did not adequately fit in any other indicator.

The six Process of Science indicators were: 1) observing or describing natural phenomena, 2) observing demonstrations of labs, 3) measuring objects, 4) experimenting and/or conducting labs, 5) displaying and organizing data, analyzing findings, and using statistics, and 6) interpreting results and conclusions. It should be noted that indicator number 1 simply requires the students to observe objects in nature, and indicator 2 simply requires the students to observe their teacher demonstrate a lab. If the student actually performed a lab or experiment, indicator number 4 was selected.

As described above, Applications of Science was not broken down into a number of indicators. Activities involving Application of Science included solving a practical problem, applying scientific principles to novel situations, and designing and carrying out experiments. In order to be considered as Application of Science, the response had to be a short explanation or longer (e.g., 1-3 sentences or more).

Two categories labeled as 'explain science' and 'explain science - reading' were included in the coding scheme that reflected discussing or explaining scientific concepts or ideas. These two categories overlap across Nature of Science, Habits of Mind, Process of Science, and Application of Science. The response type for these 'explain science' categories had to be a 'short explanation' or more. The first of the 'explain science' categories was not based on reading material, whereas the second was based on reading material. In order to be coded as 'explain science - reading', the task must have made explicit reference to a reading material. For example, the task may have stated, "Based on what you know and what you just read in the article, explain the process of photosynthesis". In contrast, if the task posed the same question as follows, "Explain the process of photosynthesis", it would have been coded as 'explain science', unless there was evidence in the task that students were required to read an article. Reading materials consisted of articles, books, research material etc. Science textbooks were not considered as reading materials.

It should be noted that all explanations were not coded into one of the 'explain science' categories. For example, if the explanation was related to the results obtained in an experiment, 'interpreting results and conclusions' was selected. Similarly, if the students were required to explain how they will carry out an experiment that they developed, 'design and modify experiments' was selected. The two 'explain science' categories were used to capture explanations that did not adequately fit into one of the other indicators subsumed under the other Process Learning Outcomes. For example, students may be asked to read an article and then provide a rationale regarding the validity of the sampling procedure used. A task

that has a question of this nature would be coded as 'bias/fairness detection' rather than 'explain science–reading' because it could be captured under 'bias/fairness detection'.

Process Learning Outcomes by Type of Activity. Table 5 provides the percentage of time that each process outcome occurred for instruction, assessment, and MSPAP test preparation activities. Although indicators of the process outcomes were coded in place of the overall Process Learning Outcomes, the table also indicates the percentage of tasks that were coded for the overall process outcome. Instruction activities tended to place most emphasis in Process of Science (59%), a similar emphasis on Nature of Science (32%) and Habits of Mind (32%), and less emphasis on Application of Science (22%). For assessment activities, Habits of Mind (32%), Process of Science (30%), and Application of Science (26%) were emphasized to a similar extent, and Nature of Science occurred a little less frequently (20%). In contrast, test preparation activities placed a similar amount of emphasis on all four process outcomes.

An investigation of the process outcome indicators shows that the test preparation activities were more aligned with MSPAP than instruction and assessment activities. For example, the majority of test preparation activities required students to apply science to a novel situation (52%) interpret results and conclusions (58%), and explain scientific concepts without the use of reading material (61%). In addition, many test preparation activities required students to conduct experiments and laboratory investigations (42%), display and organize data (30%), 'order, classify, sequence, or draw and label a diagram' (40%), and/or predict and formulate hypotheses (34%). The corresponding percentages for these indicators were much smaller for instruction and assessment, with the exception that nearly half of instruction activities required students to perform experiments or laboratory investigations (44%) and interpret results and conclusions (46%). Equally important is the finding that 20% of the assessment activities and 11% of the instructional activities did not contain a process, whereas only 1% of the MSPAP Test Preparation activities did not contain a process.

Table 5.
Process Learning Outcomes By Type of Activity

	Instruction n=1,666	Assessment n=1,174	MSPAP Test Preparation n=125
Nature of Science	32%	20%	52%
Models	6%	6%	14%
Pattern Recognition	2%	1%	6%
Predict/Formulate Hypotheses	23%	11%	34%
Bias/Fairness Detection	3%	2%	12%
Consider other's ideas	4%	5%	19%
Habits of Mind	32%	32%	56%
Order/Classify/Sequence/Draw and label	21%	23%	40%
Compare and Contrast	8%	6%	18%
Design/Modify Experiments	4%	4%	13%
Pose Scientific Questions	4%	4%	12%
Process of Science	59%	30%	67%
Observing	7%	3%	7%
Observing labs	3%	1%	2%
Measuring	14%	5%	24%
Experimenting/ Lab Investigations	44%	16%	42%
Display/Organize Data	21%	13%	30%
Interpret Results and Conclusions	46%	24%	58%
Applications of Science	22%	26%	52%
Explain Science	27%	45%	61%
Explain Science- Reading	10%	9%	22%
None	11%	20%	1%

In order to investigate which indicators occur most often within each of the process outcomes, a further discussion is provided here. As indicated in Table 5, the Nature of Science indicator, 'predict and formulate hypotheses', occurred more often than the other Nature of Science indicators for instruction (23%), assessment (11%), and MSPAP Test Preparation (34%) activities. It can also be noted that the results for all Nature of Science indicators were very similar for instruction and assessment activities. Also, as expected, the percentage of tasks coded for all of the Nature of Science Indicators was much higher for MSPAP Test Preparation activities than for instruction and assessment activities.

The Habit's of Mind indicator that occurred the most often for instruction, assessment, and MSPAP Test Preparation activities was 'ordering, classifying, sequencing, and drawing and labeling', with 21% of instructional tasks, 23% of assessment tasks, and 40% of MSPAP Test Preparation activities being coded in this category. It should be noted that this result could be due to the number of indicators subsumed in this category. For all of the Habits of Mind indicators, the results for instruction and assessment were very similar. In addition, the percentages for each Habits of Mind indicator for MSPAP Test Preparation activities was higher than for instruction and assessment activities.

The 'Process of Science' indicators that had the highest percentages were 'experimenting and/or conducting lab investigations' and 'interpreting results and conclusions'. The percentages for instruction, assessment, and MSPAP Test Preparation activities were 44%, 16%, and 42%, respectively, for 'experimenting and conducting lab investigations' and 46%, 24%, and 58%, respectively, for 'interpreting results and conclusions'. In contrast to Nature of Science and Habits of Mind, the results for Process of Science for instruction were more similar to the MSPAP Test Preparation activities than to the assessment activities. In general, this result was found for all Process of Science indicators.

Application of Science also varied across the type of activity. The percentage of tasks that contained the Application of Science process outcome for instruction and assessment were similar (22% and 26%, respectively). However, nearly half (52%) of the MSPAP Test Preparation activities contained Application of Science.

Differences across Grades. Table 6 depicts the percentage of tasks that contained each process outcome by grade. With respect to the Nature of Science outcomes, the results across grades were very similar, with one exception. Students in grade 5 used models for instruction (11%) and MSPAP Test Preparation (23%) activities slightly more often than the other grades.

The percentage of instruction and assessment activities that required students to demonstrate 'Habits of Mind' indicators was similar across grades with two exceptions. In general, for 2nd grade instruction and assessment activities there was a higher percentage of tasks that required students to order, classify, sequence objects, or draw and label a diagram than for other grades, especially for 5th and 8th grade for instruction, and 7th and 8th grade for assessment. In addition, for 5th and 8th grade instruction activities (6% and 8%, respectively) and assessment activities (7% and 8%, respectively) there was a slightly larger percentage of tasks that required students to design and modify experiments, as compared to the other grades . This result was also found for test preparation activities as 18% of 8th grade and 16% of 5th grade tasks required students to design and modify experiments, compared to 7% of 3rd grade tasks.

Table 6.
Process Learning Outcomes in Science Classroom Activities

	All Grades	2nd	3rd	4th	5th	7th	8th
Instruction							
Nature of Science							
Models	6%	6%	7%	6%	11%	4%	5%
Pattern Recognition	2%	0%	3%	1%	2%	4%	4%
Predict/Formulate Hypotheses	23%	21%	31%	23%	21%	21%	19%
Bias/Fairness Detection	3%	2%	4%	2%	6%	1%	4%
Consider other's ideas	4%	1%	4%	5%	8%	4%	6%
Habits of Mind							
Order/Classify/Sequence/Draw and label	21%	27%	22%	22%	17%	24%	16%
Compare and Contrast	8%	9%	6%	7%	6%	9%	9%
Design/Modify Experiments	4%	3%	3%	2%	6%	2%	8%
Pose Scientific Questions	4%	4%	4%	6%	3%	3%	5%
Process of Science							
Observing	7%	8%	12%	13%	9%	5%	1%
Observing labs	3%	6%	1%	4%	3%	0%	3%
Measuring	14%	9%	17%	16%	22%	8%	12%
Experimenting/ Lab Investigations	44%	41%	44%	46%	53%	41%	40%
Display/Organize Data	21%	10%	20%	16%	27%	21%	32%
Interpret Results and Conclusions	46%	39%	48%	49%	48%	44%	49%
Application of Science	22%	11%	18%	23%	23%	25%	29%
Explain Science- no reading	27%	28%	22%	33%	31%	28%	26%
Explain Science- reading	10%	7%	8%	8%	10%	14%	13%
None	11%	11%	13%	9%	5%	14%	12%
Assessment							
Nature of Science							
Models	6%	7%	6%	3%	9%	10%	2%
Pattern Recognition	1%	1%	4%	1%	1%	1%	2%
Predict/Formulate Hypotheses	11%	13%	12%	8%	13%	8%	12%
Bias/Fairness Detection	2%	0%	4%	0%	5%	1%	4%
Consider other's ideas	5%	4%	5%	3%	9%	6%	5%
Habits of Mind							
Order/Classify/Sequence/Draw and label	23%	35%	27%	24%	23%	20%	11%
Compare and Contrast	6%	8%	5%	7%	9%	9%	1%
Design/Modify Experiments	4%	2%	3%	3%	7%	2%	8%
Pose Scientific Questions	4%	3%	2%	6%	4%	2%	5%
Process of Science							
Observing	3%	4%	5%	3%	3%	0%	1%
Observing labs	1%	2%	0%	0%	1%	1%	0%
Measuring	5%	6%	5%	3%	6%	5%	5%
Experimenting/ Lab Investigations	16%	21%	19%	12%	22%	12%	12%
Display/Organize Data	13%	9%	12%	4%	20%	13%	19%
Interpret Results and Conclusions	24%	29%	26%	18%	31%	17%	23%
Application of Science	26%	19%	19%	27%	32%	27%	30%
Explain Science- no reading	45%	42%	47%	51%	52%	43%	41%
Explain Science- reading	9%	5%	4%	13%	9%	16%	10%
None	20%	16%	22%	20%	13%	23%	24%

Table 6 Continued.
Process Learning in Science Classroom Activities

	All Grades	2nd	3rd	4th	5th	7th	8th
MSPAP Test Preparation							
Nature of Science							
Models	14%	--	11%	--	23%	--	10%
Pattern Recognition	6%	--	4%	--	3%	--	10%
Predict/Formulate Hypotheses	34%	--	33%	--	36%	--	35%
Bias/Fairness Detection	12%	--	15%	--	13%	--	8%
Consider other's ideas	19%	--	20%	--	16%	--	20%
Habits of Mind							
Order/Classify/Sequence/Draw and label	40%	--	41%	--	39%	--	40%
Compare and Contrast	18%	--	20%	--	13%	--	18%
Design/Modify Experiments	13%	--	7%	--	16%	--	18%
Pose Scientific Questions	12%	--	9%	--	16%	--	13%
Process of Science							
Observing	7%	--	11%	--	3%	--	5%
Observing labs	2%	--	4%	--	0%	--	0%
Measuring	24%	--	15%	--	36%	--	28%
Experimenting/ Lab Investigations	42%	--	43%	--	52%	--	33%
Display/Organize Data	30%	--	20%	--	39%	--	35%
Interpret Results and Conclusions	58%	--	54%	--	61%	--	60%
Application of Science	52%	--	54%	--	42%	--	58%
Explain Science- no reading	61%	--	70%	--	55%	--	53%
Explain Science- reading	22%	--	22%	--	16%	--	25%
None	1%	--	0%	--	0%	--	3%

For the instruction and assessment activities, the percentage of tasks that were coded for Process of Science indicators were very similar across grades with a few exceptions. In general, the elementary and middle school on-grade (e.g., grades 3, 5, and 8) instruction activities focused on displaying and organizing data (20%, 27%, and 32%, respectively) more often than the activities in the off-grade levels (10%, 16%, and 21%, respectively for grades 2, 4, and 7). In addition, grade 5 instruction activities tended to require more experimenting and or laboratory investigations (53%) than other grades. For the assessment activities, the percentage of tasks that involved displaying and organizing data was slightly higher for on-grade levels (12%, 20%, and 19%, respectively for grades 3, 5, and 8) than off-grade levels (9%, 4%, and 13%, respectively for grades 2, 4, and 7). Similarly, the percentage of tasks that involved 'interpreting results and conclusions' was slightly higher for on-grade levels (26%, 31%, and 23%, respectively for grades 3, 5, and, 8, respectively) than off-grade levels (29%, 18%, and 17%, for grades 2, 4, and 7, respectively), except for grades 2 and 3. This result was especially true for grade 5. Finally, for test preparation activities, grade 5 tasks tended to require more experimenting and/or laboratory investigations (52%) than grade 3 (43%) and grade 8 (33%).

Regarding Application of Science, the results were very similar across grades for instruction and assessment, with the exception that there was a lower percentage of instruction tasks that were coded for application of science at the early elementary grades (i.e., grades 2 and 3). However, for MSPAP Test Preparation activities, this result was not found, as more that half (54%) of the 3rd grade MSPAP Test Preparation activities contained an application of science.

Extent to Which Activities Reflect Process Learning Outcomes. Table 7 presents a different view of how well the tasks represented the Process Learning Outcomes. This table presents the percentages of the activities for which all, part, or none of the task contained Process Learning Outcomes for the instruction, assessment and test preparation activities. As can be seen from the table, more than three-fourths of the instruction activities (80%), and nearly all of the test preparation activities (95%) reflected the Process Learning Outcomes throughout the task, compared to only about half of the assessment activities (54%). The percentages of test preparation activities reflecting the process outcomes in part or all of the task was higher than the percentages for instruction and assessment activities. Overall, 99% of test preparation activities, 89% of instruction, and 80% of assessment activities reflected the process outcomes in part or all of the task. It should be highlighted that a much larger percentage of assessment activities contain process outcomes in part of the task, as compared to instruction and test preparation activities. This indicates about one third (e.g., 26% / 80%) of the assessment activities that contain process outcomes contain the outcomes in only part of the task. In contrast, instruction and test preparation activities that contain process outcomes are more likely to contain the process outcomes throughout the entire task.

Table 7.
Extent to Which Activities Reflect Process Learning Outcomes

	Instruction (n=1,666)	Assessment (n-1,174)	MSPAP Test Preparation (n=125)
None of Task has Process Outcomes	11%	20%	1%
Part of Task has Process Outcomes	10%	26%	4%
All of Task has Process Outcomes	80%	54%	95%

Comparisons by Grade Level and Type of Activity. A repeated measures analysis of variance with one between factor (grade) and one within factor (type of activity) was conducted to determine if there were differences in proportion of activities that contain at least one process outcome. Table 8 shows an average proportion at each grade level of science activities per teacher that reflect at least one Process Learning Outcome. For this analysis, only those teachers who sent in both instruction and assessment activities were included, which represents 64% of all science teachers who sent in classroom

materials (192 out of 301)[3]. After transforming the data using an arcsin transformation, the results indicated no significant differences between grades ($F(5,186)=1.660$, $p=.146$). In addition, the interaction between grade and type of activity was not significant ($F(5,186)=.732$, $p<.600$). However, there was a significant difference within teachers for instruction versus assessment activities ($F(1,186)=16.020$, $p<.000$). This indicates that a larger proportion of teachers' instruction activities ($M=.88$) than their assessment activities ($M=.80$) reflected at least one Process Learning Outcome. This is similar to the result previously discussed that 89% of instruction compared to 80% of assessment activities contained at least one Process Learning Outcome. This provides some evidence that instruction activities may be slightly more aligned with MSPAP than assessment activities.

Table 8.
Proportion of Science Activities by Grade and Type of Activity that Contain at Least One Process Learning Outcome

	All grades (n=192)	2nd grade (n=34)	3rd grade (n=37)	4th grade (n=20)	5th grade (n=31)	7th grade (n=28)	8th grade (n=42)
Instruction activities	.88	.88	.84	.91	.93	.88	.87
Assessment activities	.80	.83	.75	.81	.85	.84	.76

Maryland Content Learning Outcomes

The classroom instruction, assessment, and MSPAP test preparation activities were also coded in terms of their content emphasis. The Maryland Content Learning Outcomes provided the basis for the classification. The MLO's related to content are life science, physical science, and earth science. Two additional categories were also added. 'General science' was reserved for those tasks that were not content specific. For example, the task may have required the student to list a set of laboratory rules, or it may have asked the student to list the steps in the scientific method. These kinds of activities are often not unique to a specific content area. 'Can not determine' was coded when the science content was not clear. For example, if the task required the students to write a report about a scientific discovery that they found in the local newspaper it may not be not clear as to which content area is the main focus.

Each activity could be coded for more than one content outcome. Thus, a content outcome had one or more opportunities to be selected for an activity. Of the 2,965 instruction, assessment, and test

[3] MSPAP test preparation activities were not included in this analysis because of small sample size. A comparison among the three types of activities (instruction, assessment, and MSPAP test preparation), would have reduced the sample to only 55 teachers, which represents only 18% of all science teachers.

preparation activities, 96% (2849) were coded solely for one content outcome. As indicated in Table 9[4], 'life science' and 'physical science' were coded the most often (39% and 35%, respectively), and 'earth science' was coded in 22% of the activities.

Table 9.
Content Learning Outcomes for Science Classroom Activities

	All grades	2nd	3rd	4th	5th	7th	8th
Life Science	39%	43%	40%	37%	32%	58%	25%
Physical Science	35%	39%	35%	28%	47%	16%	44%
Earth Science	22%	16%	26%	39%	19%	15%	25%
General Science	5%	2%	3%	1%	4%	10%	7%
Can not determine	2%	2%	1%	1%	3%	3%	4%

Differences Across Grades. Table 9 also shows the percentages for each grade level. It is interesting to note that at grades 5, 7, and 8, a dominant content area is apparent. Grades 5 and 8 tended to focus on 'physical science'. The percentage of tasks that were coded as 'physical science' was 47% for grade 5 and 44% for grade 8. However, grade 7 tended to focus on 'life science', with 58% of activities focusing on 'life science'. Such patterns were not as apparent at the lower grades, as might be expected. At grade 4, nearly equal emphasis was placed on 'life science' (37%) and 'earth science' (39%) and slightly less emphasis was focused on 'physical science' (28%). For grade 3, 40% of activities focused on 'life science' 35% focused on 'physical science', and 26% focused on 'earth science'. For grade 2, 43% of activities were 'life science' 39% focused on 'physical science' and 16% focused on 'earth science'.

Other Task Features

In addition to the process and content outcomes, other features of the classroom activities were coded including: work groups; integration with other subject areas; use of resources, manipulatives, calculators, and computers; and the type of response required by students. The analysis of instruction, assessment, and test preparation activities by these features provide additional information for describing their similarity to MSPAP tasks.

Group Work. The MLO's indicate that providing students with opportunities to work collaboratively with others is an important aspect of the learning environment. Therefore, the science classroom activities were coded as to whether they involved individual, pair, or group work. Table 10 indicates that the majority of instruction (72%) and assessment (87%), activities required only individual work. In

[4] Results are not presented by instruction, assessment, and test preparation activity because there were no differences

contrast, only half of the test preparation activities were conducted individually. It should be noted that it is possible that the percentage of actual pair or group work on these science activities is actually higher since it was coded only if it was explicitly indicated on the activity that students were to work together.

Table 10.
Type of Work Group for Science Classroom Activities

	Instruction (n=1,666)	Assessment (n=1,174)	MSPAP Test Preparation (n=125)
Individual	72%*	87%	50%
Group	12%	4%	4%
Individual and Group	16%	9%	46%

*Note- 1% of the instruction activities were actually class discussion

Integration with Other Subject Areas. The majority of the MSPAP tasks that assess science are integrated with other subject areas including mathematics, social studies, reading, and writing. Therefore, the science classroom activities were analyzed in terms of whether they were integrated with other subject areas. An activity was coded for reading integration if the student was required to read a story, article, reference book, or other type of reading material in order to solve the science questions. Reading science textbooks was not considered as reading integration. Integration with writing was selected if the activity required students to provide long explanations (e.g. a paragraph or more) or to write in their journal.

Each activity could be coded for more than one category for integration. Of the 1,666 instruction activities, 20% were coded for more than one integration category, and of the 1,174 assessment activities, 27% were coded for more than one integration category. However, 54% of the MSPAP test preparation activities were coded for more than one integration category. Table 11 indicates the percentage of times each category of integration was coded for an activity when one or more integration categories were coded. The table also indicates the percentage of times the activities required no integration.

Overall, 40% of instruction activities, 29% of the assessment activities, and 46% of the test preparation activities were integrated with other subject areas. The most common forms of integration regardless of the type of activity were within the writing area, although for instruction there was an equal emphasis placed on reading and writing. Overall, 23% of instruction activities, 28% of assessment activities, and 53% of test preparation activities involved writing integration. As an example, for the instruction activities, 18% were coded for long explanations and 6% were coded for journal. It should be noted again that since an activity could be coded for more than one type of writing, the sum of the

for type of activity.

percentages for long explanations and journal writing may be slightly higher than the overall percentage for writing.

Table 11.
Integration of other Subject Areas in Science Classroom Activities

	Instruction (n=1,666)	Assessment (n=1,174)	MSPAP Test Preparation (n=125)
Mathematics	18%	11%	32%
Social Studies	4%	4%	11%
Reading Process	23%	15%	41%
Writing	23%	28%	53%
Long Explanations	18%	26%	47%
Journal	6%	3%	9%
No Integration	60%	71%	54%

Use of Manipulatives, Resources, Calculators, and Computers. Many of the MSPAP tasks require students to use manipulatives and/or other resources. Table 12 indicates the percentage of classroom instruction, assessment, and test preparation activities that were coded for the use of manipulatives, calculators, computers, and/or reference materials. It should be noted that each activity could be coded for more than one of these categories, although this occurred infrequently. The percentage of tasks that contained no manipulatives or resources was 41% for instruction, 70% for assessment, and 30% for test preparation. Of the 59% of instruction activities that were coded for one or more manipulatives or other resources, 92% contained only one. Of the 30% of assessment activities that were coded for one or more manipulatives or other resources, 88% contained only one. Of the 70% of test preparation activities that contained one or more manipulatives or other resources, 91% contained only one.

As indicated in the table, 42% of the instruction activities, 15% of the assessment activities, and 42% of the test preparation activities required the use of manipulatives for experimentation. Other types of manipulatives, such as models or materials used for measuring, were coded less often than manipulatives used in experimentation. The use of reference materials (i.e., articles, encyclopedias, or reference books) was 9%, 8%, and 18%, respectively, for instruction, assessment, and MSPAP Test Preparation activities. Calculators were coded less frequently and evidence for the use of computers in the activities was rare.

Table 12.
Manipulatives, Calculators, Computers, and Resources Used with Science Classroom Activities

	Instruction (n=1,666)	Assessment (n=1,174)	MSPAP Test Preparation (n=125)
Manipulatives: experiment	42%	15%	42%
Manipulatives: other	10%	8%	11%
Reference materials	9%	8%	18%
Viewing Films	2%	1%	2%
Calculator	1%	0%	0%
Computer/Internet	1%	1%	2%
None	41%	70%	30%

Response Required of Student. The MSPAP tasks that are scored for science require students to respond in a variety of ways including providing short answers, short and long explanations, graphic organizers, journal entries, and graphs, tables, and charts, etc. A short answer was defined as a word or short phrase, whereas a short explanation required approximately one to three sentences. A long explanation required a paragraph or more. The science classroom activities were coded according to the response required of the student in order to examine the extent to which they reflect MSPAP response types. Each classroom activity could be coded for more than one response type. Thus, a response type had one or more opportunities to be selected for an activity. Of the 2,965 instruction, assessment, and test preparation activities, 28% (836) were coded solely for one response type, 31% were coded for two response types, 23% were coded for three response types, 11% were coded for four response types, and 5% were coded for five response types. It should be noted that approximately 3% of the activities were coded for more than five response types.

As indicated in Table 13, the 'short answer' category was coded more often than the other response types. Overall, approximately 63% of the instruction, 56% of the assessment, and 74% of test preparation activities required at least one 'short answer'. For the instruction activities, the next most frequently coded categories were 'short explanation' (51%), 'visual representation' (33%), 'making a chart, table, or graph' (22%), and 'long explanation' (18%). For the assessment activities, the next most frequently coded categories were 'short explanation' (48%), 'multiple choice, matching, true/false' (36%), 'visual representation' (30%), and the requirement for students to provide a 'long explanation' (26%). For the 3rd, 5th, and 8th grade test preparation activities, 'short explanation' was the next most frequently coded category (73%), followed by 'visual representation' (51%), 'long explanation' (47%), and 'graphic organizer' (25%).

As might be expected, the response types that were coded for the MSPAP test preparation activities, as compared to those coded for the instruction and assessment activities, were more similar to the

response types of MSPAP tasks. For example, although the top 5 response type categories that were selected for instruction, assessment, and test preparation activities included 4 of the same categories (i.e. short answer, short explanation, visual representation, and long explanation), they were ranked differently across the type of activity. For instruction and assessment activities, 'long explanations' were the fifth most common response type, whereas for test preparation activities, 'long explanation' was ranked as the third most common response type.

Table 13.
Response Type for Science Classroom Activities

	Instruction (n=1,666)	Assessment (n=1,174)	MSPAP Test Preparation (n=125)
Multiple Choice/Matching	10%	36%	6%
Ordering/Sequencing/Classifying	9%	10%	13%
Short Answer	63%	56%	74%
Short Explanation	51%	48%	73%
Long Explanation/ Research Report	18%	26%	47%
Graphic Organizer	7%	8%	25%
Journal	6%	3%	9%
Chart/ Table/ Graph	22%	7%	23%
Oral Presentation	3%	4%	6%
Class Discussion	12%	4%	16%
Visual Representation	33%	30%	51%
Other*	3%	4%	8%
None/ Can not determine	1%	1%	1%

*Other: games, creating or modifying maps, portfolios, and show your work

Similarity to MSPAP Tasks

The classroom instruction, assessment, and MSPAP test preparation activities were coded with respect to their similarity to MSPAP tasks. In particular, the process needed for solution, the type of responses required of students, and the format of the responses were considered in order to classify the activities according to one MSPAP-like level. Five levels were defined, as described below. It should be noted that in the coding scheme, the first category could be coded in conjunction with one of the other four categories. Therefore, the only possible combination of two codes is with first level and one other level. If a task received two codes, an emphasis was selected. However, the results presented here allowed for only one code on this scale, and the emphasis was used as the basis of determining the main level on the MSPAP scale.

The first level includes those activities that were considered 'Not at all like MSPAP'. This category reflects tasks that contain mostly multiple-choice, matching, true-false, or short answer formats. Thus, the first category does not require the utilization of the Process Learning Outcomes as defined by the

MLO's and MSPAP. Although some of the skills required in items of these types may also be required by MSPAP tasks, overall the tasks themselves are not considered to be similar to MSPAP tasks. Again, this level could be coded with other levels on the MSPAP scale. For example, a task that had a set of multiple-choice items along with one or more explanations would be coded in this category, as well as, in one of the other MSPAP level categories. In addition, an emphasis was selected. If the multiple-choice items were determined to be the main emphasis in the activity, then the activity would be considered as 'Not at all like MSPAP'. However, if the other aspects of the task seemed to be more dominant, the task was recoded to emphasize that particular level. If the emphasis was determined to be equal between the two components, then the task was recoded at the higher level.

The other four levels include activities that are similar to MSPAP tasks to some extent: 'MSPAP-like 1', 'MSPAP-like 2', 'MSPAP-like 3', and 'MSPAP-like 4'. 'MSPAP-like 1' tasks reflect at least one process outcome, but includes only certain processes. This is because this category reflects tasks that require no interpretation of explanation of work. Therefore the only processes that are included in this level are 'pattern recognition', 'order, classify, sequence, or draw and label', 'measuring' and 'display and organize data'. 'MSPAP-like 2' tasks reflect at least one process outcome, and usually not more than two process outcomes. These tasks have some of the elements of MSPAP tasks, but are not as extended as MSPSP tasks and do not require the same degree of the process outcomes as required by MSPAP. 'MSPAP-like 3' tasks typically reflect 2-5 process outcomes, and are similar to MSPAP in terms of the processes and response types that are required. However, they are not as extended as a MSPSP task. 'MSPAP-like 4' tasks are very similar to MSPAP because they heavily involve the process outcomes. In addition, the responses required of the students are similar to MSPAP, and the tasks are very extended and set in a realistic context. See Figure 1, which provides a description of the types of tasks that fall into each level, as well as, an example for each level.

Figure 1.
Description of the Five Levels of the Science MSPAP Scale

MSPAP Levels	Descriptors of Levels
Not at all like MSPAP	Process Outcomes: None Response Type: Selected response and short answer Integration: Rarely Manipulatives/Resources: Rarely Overall similarity to MSPAP: Not like MSPAP Example: Answer a set of multiple-choice items
MSPAP-like 1	Process Outcomes: Only one, rarely two, but the only process outcomes that may be selected are 'pattern recognition', 'order, classify, sequence, or draw and label', 'measuring' and 'display and organize data' Response Types: Ordering, sequencing, short answer, visual representation; no interpretation or explanation of work (e.g., explanations can not be included in this category) Integration: Rarely Manipulatives/Resources: Rarely Overall similarity to MSPAP: Requires at least one Process Learning Outcome, but includes only certain processes. Does not require MSPAP-like response types. Example: Classify each of the following terms as an acid or base
MSPAP-like 2	Process Outcomes: Reflect at least one, and usually not more than two Process Learning Outcomes; often contain 'explain science' Response Types: At least one short explanation or one long explanation Integration: Sometimes Manipulatives/Resources: Sometimes Overall similarity to MSPAP: Has some elements of MSPAP tasks, but to a limited extent; not as extended as MSPAP tasks, and do not contain the same the level of Process Learning Outcomes as required by MSPAP Example: Read an article and provide two short explanations
MSPAP-like 3	Process Outcomes: Usually 2-5 process outcomes are selected Response Types: Often has several short explanations (e.g., 3 or more) or 2 or more long explanations in paragraph form; charts and visual representations are also common Integration: Sometimes Manipulatives/Resources: Often Overall similarity to MSPAP: Similar to MSPAP tasks in terms of process and format, but is not as extended as a MSPAP task Example: Read an article and provide two short explanations and one long explanation Example: Conduct an experiment, and answer several short explanations
MSPAP-like 4	Process Outcomes: Usually 4-7 process outcomes are selected; must have Application of Science, with rare exceptions Response Types: Typically contain at least four short explanations or two long explanations; charts and visual representations are also common Integration: Often Manipulatives/Resources: Typically Overall similarity to MSPAP: Extremely similar to MSPAP because these tasks heavily involve the Process Learning Outcomes; response types are similar to MSPAP; the tasks are very extended and set in a realistic context Example: Read an article and provide two short explanations and one long explanation; in addition, write a letter to the editor of the journal indicating how the experiment could have been improved upon Example: Design and carry out an experiment to a research question of interest; write an article for the school newspaper illustrating your procedure, results, and conclusions

Table 14 presents the percentage of times each MSPAP-like level was coded. In general, differences across grades were negligible. However, the percentages of tasks coded for each level on the MSPAP scale differed for instruction, assessment, and test preparation activities. The most common type of instruction activity was at the 'MSPAP-like 3' level (41%), followed by 'MSPAP-like 2' (26%), 'Not at all like MSPAP' (13%), 'MSPAP-like 4' (12%), and 'MSPAP-like 1' (8%). However, for assessment activities, the percentages at each level were quite different. 'Not at all like MSPAP' was selected for 33% of assessment tasks, followed by 'MSPAP-like 2' (27%), 'MSPAP-like 3' (26%), 'MSPAP-like 4' (10%), and 'MSPAP-like 1' (4%). Therefore, the assessment activities contain a larger percentage of tasks that are not aligned with MSPAP (33%) as compared to the instruction activities (12%).

As might be expected, the 3rd, 5th, and 8th grade MSPAP test preparation activities, as compared to the instruction and assessment activities, were more similar to MSPAP tasks. The most frequently coded levels for the test preparation activities were 'MSPAP-like 4' (38%) and 'MSPAP-like 3' (36%). The next most frequently coded task type for test preparation activities was 'MSPAP-like 2' (22%). The 'MSPAP-like 1' level was selected for only 3% of the tasks, and 'Not at all like MSPAP' was selected for only 2% of the MSPAP test preparation activities.

Table 14.
MSPAP-like Levels for Science Classroom Activities -- For Each Grade

		Grade					
	All grades	2nd	3rd	4th	5th	7th	8th
Instruction							
Not at all like MSPAP	13%	11%	16%	11%	6%	19%	14%
MSPAP-like Levels							
MSPAP-like 1	8%	14%	7%	4%	6%	8%	7%
MSPAP-like 2	26%	32%	22%	27%	32%	21%	24%
MSPAP-like 3	41%	34%	44%	47%	41%	39%	40%
MSPAP-like 4	12%	8%	11%	12%	16%	13%	15%
Assessment							
Not at all like MSPAP	33%	25%	34%	34%	28%	38%	38%
MSPAP-like Levels							
MSPAP-like 1	4%	8%	5%	1%	3%	4%	2%
MSPAP-like 2	27%	32%	30%	32%	29%	21%	23%
MSPAP-like 3	26%	26%	23%	25%	24%	26%	30%
MSPAP-like 4	10%	10%	8%	7%	17%	12%	7%
MSPAP Test Preparation							
Not at all like MSPAP	2%	--	2%	--	0%	--	3%
MSPAP-like Levels							
MSPAP-like 1	3%	--	0%	--	7%	--	5%
MSPAP-like 2	22%	--	22%	--	19%	--	23%
MSPAP-like 3	36%	--	39%	--	39%	--	30%
MSPAP-like 4	38%	--	37%	--	36%	--	40%

Comparisons by grade level and type of activity. To determine if there were differences in the level of similarity to MSPAP tasks between grade level and within instruction and assessment activities, a repeated measures analysis of variance with one between factor (grade) and one within factor (type of activity) was conducted. Table 15 shows an average MSPAP level for teachers at each grade level.[5] For this analysis, only those teachers who sent in both instruction and assessment activities were included, which represents 64% of all science teachers who sent in classroom materials (192 out of 301). Similar to the previous repeated measures analysis for the Process Learning Outcomes, the MSPAP Test Preparation activities were not included in this analysis because of small sample size. The results indicated that no significant difference was found between grade levels ($F(5,186)=1.793$, $p=.116$). In addition, the interaction between grade and type of activity was not significant ($F(5,186)=.386$, $p=.777$). However, a significant difference within teachers for instruction versus assessment activities ($F(1,186)=36.755$, $p<.000$) was found. This indicates that the mean MSPAP level for instruction activities ($M=3.23$) was higher than the mean MSPAP level for assessment activities ($M=2.78$). This result is consistent with the finding from the previous repeated measures analysis that a larger proportion of teachers' instruction activities than assessment activities reflected at least one Process Learning Outcome. These results along with the results presented in Table 14 provide some evidence that instruction activities are slightly more aligned with MSPAP than assessment activities.

Table 15.
Similarity to Actual MSPAP Science Tasks by Grade and Type of Activity

	All grades (n=192)	2nd grade (n=34)	3rd grade (n=37)	4th grade (n=20)	5th grade (n=31)	7th grade (n=28)	8th grade (n=42)
Instruction activities	3.23	3.01	3.04	3.34	3.47	3.33	3.28
Assessment activities	2.78	2.83	2.61	2.76	2.97	2.93	2.68

Discussion

This paper presented the results of analyses investigating the similarity of classroom materials collected form science teachers in Maryland to the Maryland Learning Outcomes and MSPAP. A coding scheme was developed to reflect important features of tasks including the extent to which the activities reflected the Maryland Learning Outcomes for science, their overall similarity to MSPAP-like tasks, the

[5] The MSPAP scale has 5 levels where 'Not at all like MSPAP' received a code of 1, and "MSPAP-like 4' received a code of 5.

response required of the student, integration with other content areas, the use of manipulatives and other resources, and the incorporation of group work.

The Maryland Learning Outcomes are divided into Process Learning Outcomes and Content Leaning Outcomes. The Process Learning Outcomes for science are Nature of Science, Habits of Mind, Process of Science, and Applications of Science. However for this analysis, the first three process outcomes were broken down into more specific indicators which provided a more detailed analysis of the processes that students were required to do. Applications of Science remained intact as defined by the MLO's. Two additional categories were added to the coding scheme that were labeled as 'explain science' and 'explain science – reading'. These categories were added because each of the of the Process Learning Outcomes required students to explain scientific phenomena. Because they overlapped with each of the other process outcomes, they were used to capture explanations that could not be adequately represented elsewhere. The category 'explain science' was used to represent explanations that were not based on reading material, and the 'explain science – reading' category was used to represent explanations that were based on reading material, such as articles.

Tasks were analyzed with respect to whether they contained Process Learning Outcomes throughout the task, in part of the task, or in none of the task. In general, a large percentage of tasks contained process outcomes throughout the task. This was especially true for instruction (80%) and test preparation activities (95%). Assessment activities, as compared to instruction and test preparation activities, had a lower percentage of tasks that contained process outcomes throughout the task (54%). A significantly larger proportion of teachers' instruction activities, as compared to their assessment activities, reflected at least one Process Learning Outcome. This provides some evidence that their instruction activities were slightly more aligned with the MLO's and MSPAP than their assessment activities. However, the percentage of assessment activities that contained at least one Process Learning Outcome was also quite high.

As described above, the specific processes were not coded for Nature of Science, Habits of Mind, and Process of Science, but rather indicators of the processes were coded. However, in order to capture the larger picture about the nature of the activities in the science classroom, it was determined that the overarching process outcomes needed to be presented. Overall across instruction, assessment, and test preparation activities, it was found that Process of Science (48%) occurred more often than Nature of Science (28%), Habits of Mind (33%), and Application of Science (24%). This result was also found when isolating the classroom materials by type of activity (e.g., instruction, assessment, and test preparation activity) for instruction and test preparation activities. Instruction activities tended to place a similar emphasis on Nature of Science and Habits of Mind, and less emphasis on Application of Science.

In contrast, test preparation activities placed a similar amount of emphasis on the remaining process outcomes. For assessment activities, Habits of Mind, Process of Science, and Application of Science were emphasized approximately equally, and Nature of Science occurred a little less frequently.

With respect to Content Learning Outcomes, there were some differences in the percentages of activities reflecting the different content areas across grades. More specifically, grades 5 and 8 tended to focus on 'physical science', and grade 7 tended to focus on 'physical science'. No such patterns were found at the lower grade levels.

Other task features were examined as well. Manipulatives and other resources were used in instruction and test preparation more often than for assessment. This result is reasonable because manipulatives for experiments were coded more often than other types of resources, and instruction and test preparation activities seem more conducive to experimentation than assessment. Regarding group work, activities that required both individual and group work were not commonly seen in the instruction or assessment activities, but were more common in the test preparation activities. This could be due in part to the fact that teachers may not have specified on the instruction and assessment activities that group work was part of the activity. If this specification was not made in writing, the task was treated as requiring individual work only.

With respect to integration across other subject areas, more of the MSPAP test preparation activities required integration than both the instruction and assessment activities, with more of the instruction than assessment activities requiring integration of some sort. Writing integration was the most common form of integration across the type of activity, as might be expected in the science subject area, but the percentage of test preparation activities that were integrated with writing was much higher than the percentage of instruction or assessment activities that required writing integration. It should be noted that for instruction, reading and writing were emphasized equally. Assessment activities also did not utilize reading integration nearly as much as instruction or test preparation activities.

The response types that were coded in test preparation activities were more similar to MSPAP response types, as compared to instruction and assessment. For example, 3 out of 4 test preparation activities contained at least one short explanation, compared to 1 out of 2 instruction and assessment activities. Furthermore, 1 out of 2 test preparation activities required the students to write long explanations, compared to only 1 out of 4 assessment activities and 1 out of 5 instruction activities. Finally, graphic organizers were emphasized in 1 out of 4 test preparation activities. The corresponding percentages for instruction and assessment activities were much smaller.

In further considering the relationship between the classroom activities and the Maryland Outcomes, an important comparison can be made between the overall similarity of the activity to the tasks on the

MSPAP. This comparison yields the most encompassing measure that was used to relate the classroom activities to the Maryland Learning Outcomes and the MSPAP tasks. Overall, it was found that, only 5% of the MSPAP test preparation activities fell into the two lowest categories on this scale. Further, 38% of the test preparation activities fell into the highest MSPAP-like category, and approximately 3 out of 4 test preparation activities were in one of the highest two levels. Instruction and assessment activities showed different patterns. In general, instruction activities were more similar to MSPAP than assessment. Although, 13% of the instruction activities fell at the lowest MSPAP-like level, more than half of the instruction activities received one of the highest two levels on the MSPAP scale. For assessment activities, approximately one third of the activities were are the lowest level and another third were in the highest two levels.

Many of these findings suggest that the instruction, assessment, and test preparation activities sent in by teachers in the state of Maryland during the 1997-98 school year reflect the Maryland Learning Outcomes from a moderate to a high degree. However, instruction, and especially assessment activities are less similar to the Maryland Learning Outcomes and MSPAP, than the test preparation activities.

Comparisons across Mathematics, Reading, Writing, Social Studies, and Science

This research program has examined classroom materials across the five subject areas of mathematics, reading, writing, social studies, and science. This allows some comparisons to be made across the subject areas and to examine any trends that may exist across the different subject areas.

Table 16 indicates the percentages of instruction, assessment and test preparation activities that were coded at the MSPAP-like 3 and MSPAP-like 4 levels across the different subject areas. The upper most levels of the MSPAP scale were defined very similarly across all subject areas. However, the other levels of the MSPAP scale are not presented because the interpretation of the lower levels is substantively different across the subject areas. For instruction activities, science (53%) had the highest percentage of tasks that were coded into the highest two levels of the MSPAP scale, followed by writing (42%), reading (30%), social studies (29%), and mathematics (14%). A similar pattern was found for assessment activities. However, the percentage of tasks that were at the two highest levels were more similar for science (36%), reading (36%), and writing (35%) than they were for instruction. The social studies subject area yielded 29% of the tasks at the highest levels, and the mathematics subject area only had 10% of assessment activities in those levels. For test preparation activities, science materials were at the highest 2 levels in 74% of tasks, followed by writing (66%), social studies (64%), mathematics (52%), and reading (49%).

Table 16.
MSPAP-like Levels for Each Subject Area by Type of Activity

	Instruction	Assessment	MSPAP Test Preparation
Mathematics	(n=1940)	(n=1388)	(n=125)
MSPAP-like 3	9%	6%	22%
MSPAP-like 4	5%	4%	30%
Reading	(n=1411)	(n=906)	(n=85)
MSPAP-like 3	22%	25%	24%
MSPAP-like 4	8%	11%	25%
Writing*	(n=1101)	(n=742)	(n=80)
MSPAP-like 3	28%	22%	28%
MSPAP-like 4	14%	13%	38%
Social Studies	(n=1433)	(n=1095)	(n=103)
MSPAP-like 3	20%	21%	25%
MSPAP-like 4	9%	8%	39%
Science	(n=1,666)	(n=1,174)	(n=125)
MSPAP-like 3	41%	26%	36%
MSPAP-like 4	12%	10%	38%

*MSPAP-like 3 reflects an MLO and some stages of writing process, whereas MSPAP-like 4 reflects an MLO and all stages of writing process

In examining the differences across the subject areas it may seem that science materials are very aligned with the MLO's and MSPAP, whereas mathematics materials are not. However, the nature of the subject area yields different opportunities for the process outcomes to be demonstrated for at least two reasons. First, explanations are a very important component of the MLO's, and some may argue that mathematics can not be expected to display the same level of explanations as other subject areas where writing plays a more significant role in the learning process. However, this is not to say that the incorporation of more explanations and/or justifications can not be demonstrated in the mathematics classroom. The evidence provided from this project suggests the contrary. Second, the process outcomes for some subject areas delineate processes that have traditionally be incorporated into classroom instruction and assessment practices. For example, the MLO's for science contain the outcome, Process of Science, which focuses on experimenting, conducting laboratory investigations, displaying and organizing data, and interpreting results and conclusions. These activities have traditionally been conducted in the science classroom, so it may not be surprising that they are expressed to a high degree in the science activities.

Another important consideration is with the MLO's themselves. For example, the Process Learning Outcomes for social studies were defined in detail with more apparent boundaries, than the Process Learning Outcomes for science. As an example, in order to obtain the indicator 'obtain and use textual

information' in the Skills and Processes social studies learning outcome, 5th grade students were required to "obtain, interpret, organize, and use information from reading, asking questions, observing and listening". The coding scheme for social studies was developed to model the specifications in the MLO's. Therefore, if the student was reading textual information, and simply pulling information from the text, 'explain social studies – reading' was chosen in place of "obtain and use textual information" because the students in this example are not required to interpret and/or organize information.

By examining each of the subject areas, important differences across subject areas can be examined. However, differences that may exist across subject areas need to be considered and interpreted with caution. This research program indicated that there is some evidence that all the subject areas have a large percentage of test preparation activities that are aligned with the MLO's and MSPAP. In contrast, in all subject areas, the extent to which instruction and assessment activities are aligned with the MLO's and MSPAP is more limited. The classroom materials used in this study were only collected at one time period, and no baseline data is available regarding the extent to which classroom materials were aligned with the MLO's and MSPAP prior to the inception of MSPAP. However, MSPAP began in 1992, and the materials used in this study were collected during the 1996-97 and 1997-98 school years, so at least 4-5 years has passed since the inception of MSPAP when these materials were collected. At this time point, evidence shows that the majority of the instruction and assessment activities reflect some aspects of the MLO's and MSPAP, but not to the degree to which they could.

References

Maryland State Board of Education. (1995). Maryland school performance report: State and school systems. Baltimore, Maryland.

National Research Council. (1996). National Science Education Standards. Washington, DC.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☑ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").