Evidence for the validity and reliability of environment based classroom assessments as measures of the Washington State essential academic learning requirements[1]

Pacific Education Institute
Technical Report Number 7

Catherine Taylor, University of Washington
Kathryn Kurtz Smith, Washington Environmental Educational Assessment Project
Margaret Tudor, Washington State Department of Fish and Wildlife
Lynne Ferguson, Washington Forest Protection Association
Oksana Bartosh, University of British Columbia

[1] The authors gratefully acknowledge the public and private agencies who funded this research including: National Environmental Education Training Foundation (NEETF), National Project Learning Tree, Discuren Foundation, Washington Mutual, and the Environmental Education Training Partnership (EETAP); the Washington Forest Protection Association, the Washington Department of Fish and Wildlife, the University of Washington, National Project Learning Tree and the State Environment and Education Roundtable contributed through in-kind donations.

Table of Contents

Table of Tables

Abstract

Criterion referenced, integrated assessments matching the formats of items for the *Washington Assessment of Student Learning* (WASL) were developed as indirect measures of the Pacific Education Institute's (PEI) benchmark performances for three strands of environmental literacy: research-based inquiry, civic participation, and systems analysis. The purpose of the study presented in this technical report was to obtain evidence for the reliability and validity of scores from the integrated WASL-like tasks. For the purposes of the study, scores were obtained for each PEI strand and for the PEI WASL-like assessments as a whole. Each WASL-like assessment was completed by at least 100 students. All items were scored by three raters. The score assigned to an item was the modal score. In the event that three raters could not agree on a score, a fourth scorer – and expert on the PEI assessments – resolved discrepancies. Analyses for inter-rater reliability and correlations between item scores and scores on the discipline-based (WASL) were conducted to evaluate item quality. Most items had acceptable reliability and validity data. Items that did not function well were eliminated from the strand and total scores. Alpha coefficients and correlations between the PEI WASL-like assessment scores and scores on the WASL were obtained as evidence for the reliability and validity of scores. As predicted, these integrated WASL-like test and task scores have moderate to strong correlations with student scores on the math, writing and reading WASL. The results of the study indicate a solid relationship between performance on integrated environment-based assessments and student achievement on WASL.

Background Information

PEI is dedicated to the development of an environmentally literate citizenry and to provide assessment and evaluation tools (criteria and evidence) that will ensure quality implementation of programs, curriculum and educational practice. Instructional practices that build environmental literacy focus on four areas of development: 1) conceptual understanding, 2) cognitive processes; 3) procedural skills; and 4) affective processes for the purposes of a) deepening personal, socio-cultural, and environmental understanding related to the interplay between natural and social systems and b) working toward the development of sustainable societies, cultures and environments. These instructional practices are integral to established environmental education programs whose goals are generally to promote, awareness, appreciation, knowledge, and stewardship of natural resources, plus seek commitment to address environmental issues through informed decisions, responsible behavior, and constructive actions (Council for Environmental Education, 2002; American Forest Foundation, 2003; The Watercourse and the Council on Environmental Education, 1999).

A meta-analysis of research by environmental educators regarding environmental literacy (Volk and McBeth, 1998) resulted in the following conclusions about the assessments used in environmental education research: 1) "cognitive skills" were missing from the assessments; 2) studies across the country were haphazard in their attention to the reliability and validity of the assessment tools used (32% of the research projects reported evidence for either validity or reliability but not both, and 34% did not report evidence for either reliability or validity); 3) research in the field came primarily out of four states, all of which were near the Great Lakes region; and 4) the research tools used were not comparable from one study to another, which

prevented the comparisons needed to develop a national picture of how the populace is progressing towards environmental literacy.

Environmental education programs must be considered useful in aiding student achievement if they are to be included in the standard public school curriculum. With the introduction of the curriculum standards movement and more recently the passage of No Child Left Behind (NCLB), schools are struggling to meet accountability requirements in reading, math, and science. Educators are reluctant to add untested subjects or to address integrated applications such as environmental education curricula. Even with documentation to demonstrate that environment-based education improves student motivation and achievement levels (Lieberman & Hoody, 1996), administrators are hesitant to burden their educators when environmental education is considered "more to do." The NCLB legislation attaches rewards and sanctions to schools based on whether students in these schools are making adequate yearly progress (AYP) as defined by performance on the state assessments. When school funding is tied to test scores, educators narrow the curriculum to tested content in order to ensure success on the tests used to judge students and schools (Smith, 1991).

Environmental education programs have responded by correlating instructional materials to State and National discipline-based standards (e.g., *Curriculum and Evaluation Standards for School Mathematics* developed by the National Council for Teacher of Mathematics). Environmental education instructional materials involve integration of reading, writing, mathematics, science, economics, geography, and civics. Environmental literacy goals in environmental education require students to take knowledge from the social sciences, language arts, mathematics, and natural sciences, synthesize that knowledge and apply it in authentic contexts. Learning a traditional subject matter in an environmental context has been shown to increase critical and creative thinking skills that contribute to higher levels of achievement in the subject disciplines (Lieberman & Hoody, 1998; Lieberman, Hoody & Lieberman, 2000) . The instructional strategies characteristic of environmental education programs also contribute to higher student achievement (Marzano, Pickering & Pollock, 2001). While different studies have supported these claims (National Environmental Education and Training Foundation, 2002) environmental education remains an optional rather than central aspect of education in the United States.

In 1998, a group of environmental educators in Washington State set out to contribute to the body of research that supports the effectiveness of environmental education strategies. The Environmental Education Assessment Project (EEAP – now PEI)[2] began with two important goals: to identify the valued performances that demonstrate environmental literacy (benchmark performances) and to align environmental education goals with thinking skills, the state curriculum standards and the state tests. These goals required creating performance assessments and standardized tests that measure state standards through environmental contexts. It was hoped that, doing this, environment based education programs known to increase student motivation would have tools to measure student achievement in the content knowledge and thinking processes measured on the state tests. In addition, if the state curriculum standards contained standards that focused on thinking processes and integrating information to apply it to new situations, then the environmental education curriculum materials should include classroom

[2] Beginning in 2004, PEI work was coordinated by the Pacific Education Institute, a consortium of public and private organizations with concerns about environmental literacy and stewardship and with vested interests or legal mandates to contribute to both. In future documents, all PEI work will be referred to as PEI work.

assessments wherein students apply the same knowledge and skills measured by state assessments in meaningful contexts.

The PEI benchmark performances require integration of multiple layers of information, thinking processes, and understandings. A benchmark performance is an overall description of what a student should know and be able to do and contains the steps a student would need to take to demonstrate an acceptable level of performance (Taylor, 2002). Critical and creative thinking processes, subject matter knowledge, and procedural skills form the steps of these benchmark performances (See Appendix A for an example benchmark performance description).

Representatives from nineteen non-formal environmental education programs, 13 colleges and universities, and 29 formal K-12 education programs were involved in the initial process of developing the PEI benchmark performances. These educators began their work by answering the question, "What valued work can students do when they have been educated through environmental education programs?" Using the backwards design model (Wiggins & McTighe, 1999, pp. 7-19), the educators established exit level performances for high school students for four strands of competency within the domain of environmental literacy: system analysis, research-based inquiry, civic participation, and understanding of and expression in the language, visual and performing arts. Next, developmentally appropriate versions of the exit performances were described for two benchmark levels of school – the end of elementary school and the end of middle school or junior high school.

The assessment of these benchmark performances support the notion that assessment should be used "not only to determine what people know, but also to assess how, when and whether they use what they know" (National Research Council, 2001). The National Research Council states, "[A]ssessments must expand to encompass issues involving the organization and processing of knowledge, including participatory practices that support knowing and understanding and the embedding of knowledge in social context." (2001)

Since 1997, a strong focus in Washington State has been on scores on the *Washington Assessment of Student Learning* (WASL), Washington's state tests. The coordinators of the PEI work determined that a bridge was needed between the desired benchmark performances and the current state tests. Using the benchmarks as a guide, PEI developed paper-pencil assessments for classroom use that bridged the PEI benchmark performances directly to the state tests. The development of the "WASL-like" tasks began. (See Appendix B for an example WASL-like task.)

The WASL-like assessments are criterion referenced, integrated measurement tools that match the WASL items in format and in types of items. They differ only in their use of environmental scenarios as the focus of the assessment of reading, writing, mathematics, science, and social science knowledge and skills. Each scenario is organized around one of three PEI benchmark performances for three strands of environmental literacy: research-based inquiry, civic participation, and systems analysis.

Before using the WASL-like tasks in research projects to investigate the impact of environmental education programs on student learning, PEI needed to establish the reliability and validity of using integrated measures to assess knowledge, skills, and thinking in the traditional subject disciplines. It was also critical to members of the PEI that the integrated measures not only test for student knowledge in discreet disciplines, but that they assess students' thinking processes including problem solving, decision making, experimental inquiry skills, field investigation skills, and analysis of information. PEI wanted to maintain five characteristics in the WASL-like assessments: 1) demonstration of subject area knowledge and

skills, 2) demonstration of valued thinking processes; 3) criterion-based scoring guides; 4) focus on valued learning for environmental literacy; and 5) format to match the state test.

The case for developing critical and creative thinking skills rather than memorization of facts has long been a goal of education (Sternberg & Swerling, 1996). Many educators consider norm referenced tests poor measures of students' abilities in decision-making, problem-solving and other critical and creative thinking processes. (Resnick & Peterson, 1991; Shepard, 1989)

One way that Washington's state tests (WASL) assess thinking and problem-solving skills is through the use of performance items (e.g., short-answer, extended response, and essay items). Use of performance items requires the educator to establish scoring rules by which students' responses will be evaluated to determine the extent to which they show evidence that some or all of the performance criteria in the scoring rules are met (Taylor & Nolen, 2005). This allows the educator to determine what information, thinking processes and procedural skills the students have learned and still need to learn. Performance items can also be used to demonstrate competencies on curriculum standards in the disciplines. PEI determined that the majority of the items on the WASL-like assessments would be performance items so that there would be optimum opportunity for students to show their ability to apply disciplinary knowledge and thinking skills to solve problems, make decisions, analyze information, and communicate their understanding.

Performance items pose some problems for psychometricians and policymakers. First, the scoring of performance items requires human judgment, is time intensive, and can be expensive. Developing good rubrics goes a long way to ease the problems associated with scoring. Personnel from the Office of the Superintendent of Public Instruction (OSPI) have worked with the WASL-like tasks to ensure that they remain close to the WASL test items in format and scoring rubrics. For example, Figure 1 is a released item from the WASL reading assessment. Figure 2 is a WASL-like item from a civics task. Note that both items require students to read text and answer questions about the information provided in the text. Items also have similar formats.

The second potential problem with performance items is multi-dimensionality. This is a quality of a test item if it requires more than one cognitive dimension. One student might have strong knowledge or skills in the first dimension and weaker knowledge or skill in the second dimension. Another student might have weaker knowledge or skill in the first dimension of performance and have strength in the second dimension. Both students could earn the same item score, but the scores would have different meanings. Hence, scoring rules for performance items must be developed to allow the educator to evaluate students' different strengths and needs.

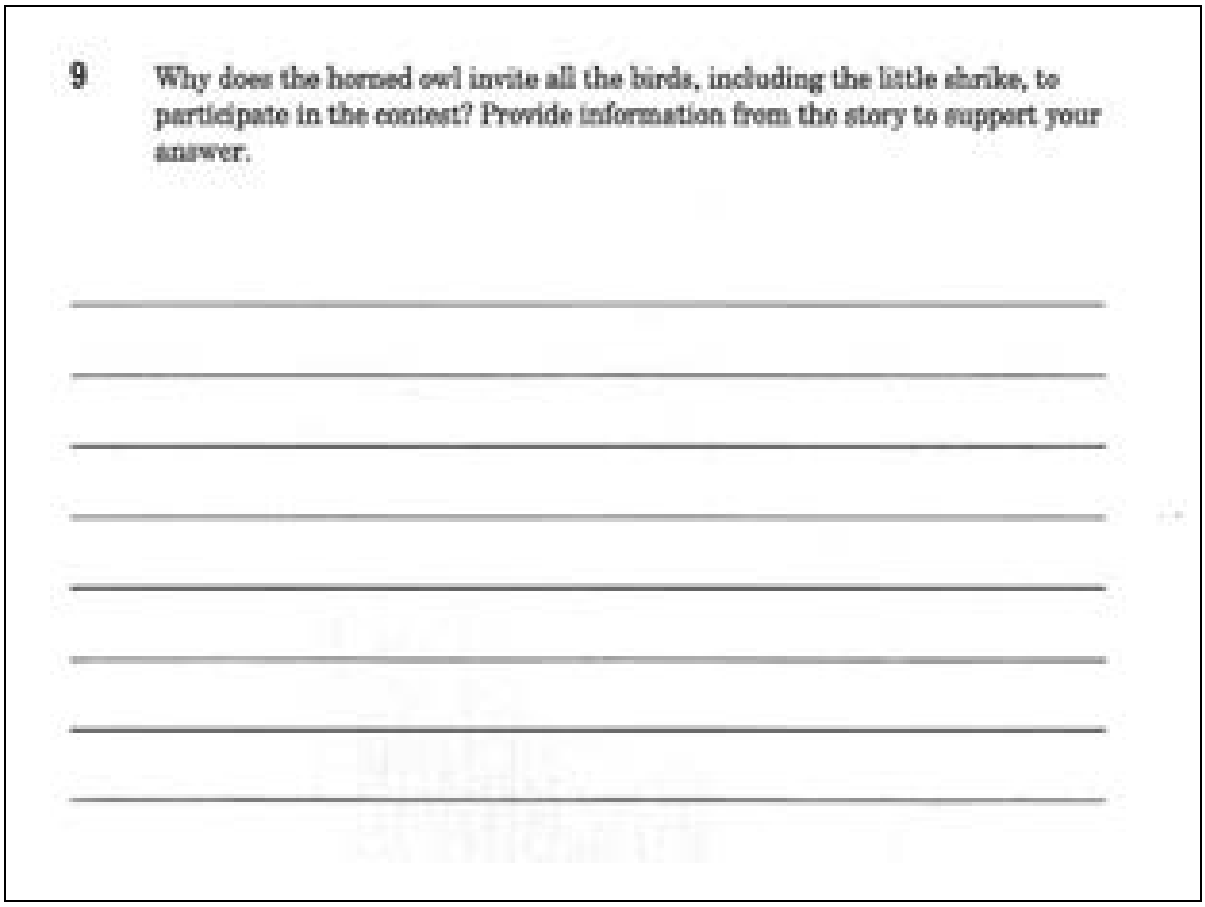


Figure 1: Example WASL Grade 7 Reading Comprehension Item

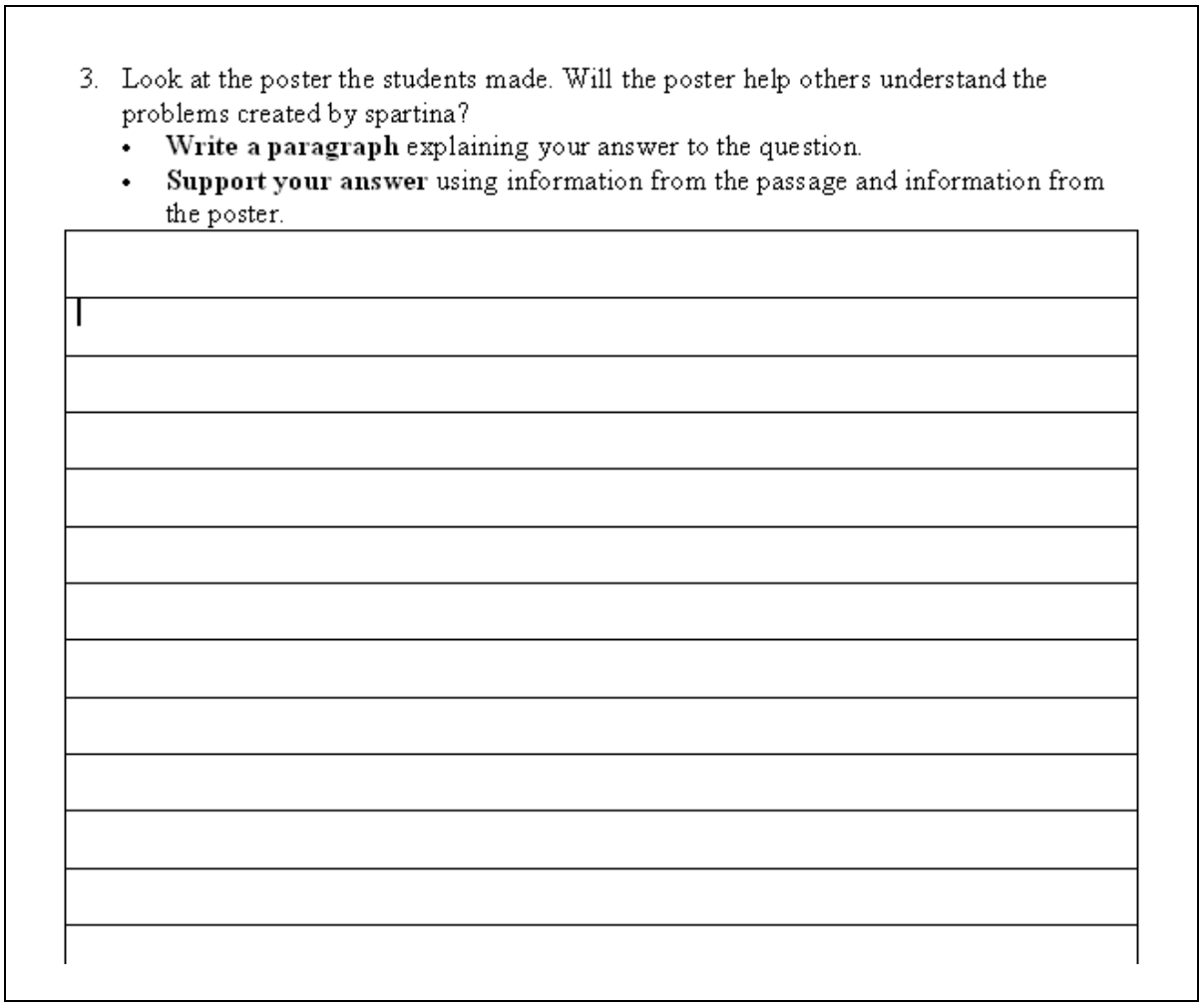


Figure 2: Example WASL-like Item Measuring Reading Comprehension

While Washington's state tests are discipline specific in their focus, items in all subject areas that require students to construct responses tap into students' skill in organizing and representing information – a set of skills needed for proficient writing. Items in science, reading, or the social sciences that require students to read data presented in graphs, charts and tables also tap into students' mathematics skills (i.e., statistics – data displays). Therefore, since the PEI WASL-like assessments required students to integrate their knowledge and skills across subject areas, the WASL-like items were likely to be multi-dimensional or multi-trait items. To address this issue, items were classified according to what knowledge and skills they tapped. Every effort was made to ensue that PEI WASL-like items met the same rigorous standards for high quality assessments that were followed in the development of WASL items.

In the rest of this paper, we describe the research that was conducted to gather evidence for the validity and reliability of the WASL-like assessments. Evidence for reliability was obtained through data on inter-rater agreement and internal consistency. Evidence for validity was obtained by correlating the items and total scores from the WASL-like assessments with scores from the *Washington Assessment of Student Learning* (WASL) tests in Reading, Mathematics, and Writing.

Research Methodology

This research looked at the relationship between the WASL-like item and task scores and the scores on the *Washington Assessment of Student Learning* (WASL) in Reading, Mathematics, and Writing are criterion-referenced achievement tests that are designed to measure whether students are achieving the Washington State Essential Academic Learning Requirements (EALRs). Scores from the WASL were used to obtain evidence for the validity of scores from the PEI WASL-Like assessments. In order to trust validity evidence, evidence is needed regarding the reliability of scores. Therefore, a second focus of the research was on analysis of reliability data.

The WASL-like assessments were developed and refined from 1998 to 2003. At least 2 WASL-like tasks were developed for each of the benchmark performance strands: systems, inquiry and civics for 3 grade levels. In other words, the benchmark performances were task analyzed into their component parts and these parts became the shells for test items. For example, the parts of the inquiry benchmark performance included: generating a research question, conducting background research related to the question, generating an hypothesis, designing an investigation, collecting data conducting the investigation, organizing and analyzing the data, and drawing conclusions. Each of these component parts became a possible focus for a WASL-like test item.

The WASL-like assessments measure student achievement in geography, civics, economics, science, reading, writing, health, and mathematics as defined the by Washington State EALRs. Each assessment was pilot tested in at least 2 classrooms. Student work was gathered to help improve the directions for each of the items and to improve the scoring rules (rubrics). The pilot testing was conducted in November of 2001. Tests and rubrics were adjusted according to the results, including improvements in directions to students, scoring rubrics, and formatting.

Study Participants

Using the improved assessments, a research study assessments was conducted in March 2003. Twenty elementary and 10 middle schools were selected for the study. Thirteen elementary and nine middle schools participated in the study and six elementary and four middle schools (eight classrooms) completed all aspects of the study. All participants were students in the K-12 public school system in the spring semester of grade 5 or grade 8. Students represented rural, suburban and urban populations in Washington State as well as geographically diverse regions of the state. Table 1 gives the number of students at each grade level who completed each WASL-like task.

Table 1
Number of Students That Completed Each WASL-like Task

Grade	Form	Strand	Number of Students
5	A	Research-Based Inquiry	138
	A	Systems	133
5	B	Research-Based Inquiry	171
	B	Systems	119
	B	Civic Participation	181
8	A	Research-Based Inquiry	197
	A	Systems	212
	A	Civic Participation	193
8	B	Research-Based Inquiry	170
	B	Systems	176
	B	Civic Participation	174

Assessment Tools

PEI WASL-like Assessments

A WASL-like task is composed of a scenario and a set of items asking students to analyze the information presented in the scenario to answer questions, propose solutions, graph data, identify information, etc. A set of three tasks (one from each benchmark strand) constitutes a test. Each task was administered on a different day.

Scores from the items within each task were classified by discipline (reading, mathematics, geography, economics, etc.) and by performance strand (research-based inquiry, systems analysis, and civic participation).

Multiple-choice items from the WASL-like assessments were scored using an answer key. Short-answer and extended response items were scored using rubrics similar to those used for the operational WASL assessments. The WASL-like assessments were scored by trained teachers.

Scoring Student Responses to PEI WASL-like Assessments

Scoring training followed the same the same training strategies used by NCS-Pearson, Washington State's scoring contractor (see Appendix B for a sample item and its scoring rubric). Prior to scoring training, PEI coordinators selected student work to represent each score point on a scoring rubric. Three student responses were selected for each score point. These pre-scored responses are called 'anchor papers.'

Volunteer teachers attended a scoring session. First, they completed the task they were to score. Then they reviewed the scoring rubric for the items within the task. They discussed the scoring rubric and examined the anchor papers. Next, the teachers practiced scoring student work that had been previously scored by expert scorers. They compared their scores with the criterion scores and discussed any discrepancies. The goal was for all teachers to consistently apply the scoring rubrics to students' responses.

Three trained teachers (raters) scored each student's responses. The student's score on an item was the most common score of the raters. If raters' item-level scores were discrepant by more than one point, a fourth rater, an expert on the PEI WASL-like tasks, scored student work to ensure an accurate score. In the case of a significant discrepancy, the expert's score became the item score for the individual student. This process was the same as that used by the State's scoring contractor (Office of the Superintendent of Public Instruction, 2002)

Washington Assessment of Student Learning (WASL)

The WASL is a set of criterion-referenced tests that assess students' achievement of Washington State's Essential Academic Learning Requirements (EALRs). The reading, listening, and mathematics tests are composed of multiple-choice, short-answer, and extended response items. The writing test is composed of two extended essay items. Students have as long as they need to complete each test. Scores from the test include strand scores (e.g., geometric sense, measurement) and total scores in Reading, Listening, Mathematics, and Writing.

Test development for the WASL tests began in the summer of 1994. Washington teachers developed models for items and performance tasks in reading, writing, listening, speaking, and mathematics. Using these models, teachers wrote prototype assessments that ranged from multiple-choice items to multiple day projects. These prototype assessments were piloted in over 80 districts throughout Washington State. Results of the pilots were used to determine the types of items that would be included on the state standardized test.

In the summer of 1995, development of the fourth grade WASL tests in Reading, Mathematics, Writing, and Listening began. Supervised by a testing contractor, Washington teachers reviewed WASL test and item specifications and modified these as appropriate to measure the EALRs. Items were written to specifications and reviewed by content review committees and a bias review committee. Fourth grade items were piloted throughout the state in the spring of 1996. Content review committees reviewed the item analysis data and selected the items to be included in the item pool for future tests. Operational testing of fourth graders began in the spring of 1997.

These same procedures were followed in the development of the tests for seventh and tenth grades. Each year of operational testing, studies are conducted to evaluate the reliability and validity of WASL scores. Technical reports are prepared and posted on the website for the Washington State Office of the Superintendent of Public Instruction (www.k12.wa.us). The technical reports present inter-rater reliability data, internal consistency reliability data, item analysis data, the results of validity studies, and the performance of students on each test. Student performances are reported by gender, ethnicity, categorical program, and for the state as a whole.

The inter-rater reliability data for WASL presents the percent of times raters give exactly the same score to students' responses to open-ended items. Exact agreement generally ranges from 70-90% for the Reading, Listening, Mathematics, and Writing items. Inter-rater reliability is also measured by computing correlations between first and second readers at the total score

level. Correlations generally range from .95 to .99 for Reading, Mathematics, and Writing. Internal consistency measures of reliability range from .88 to .89 in Reading, from .90 to .92 in Mathematics and from .79 to .85 in Writing.

Validity studies presented in the technical reports include: correlations between scores on WASL tests and scores on the *Iowa Tests of Basic Skills* (ITBS), factor analyses of WASL strand scores and ITBS subtest scores, and multiple-regressions showing how student variables such as "how far I plan to go in school" and "mother's educational level" predict WASL scores. The results of all of the studies provide solid support for the validity of scores for WASL tests.

Procedures

For the current study, each student was assigned an identification (ID) number. The student ID number was used on the WASL-like task booklets and was on the WASL and ITBS scores sent to the PEI project staff. The PEI project staff did not collect any names in association with student work or test scores.

Each participating teacher administered three, one-hour WASL-like tasks between February 15th and March 15th of 2003. Teachers received written instructions (directions for administration) on how to administer the test. Directions for Administration included oral directions for completing the demographic information on the test booklets, oral directions for completing the tasks, and directions for return of the materials.

Once students had completed the tasks, teachers returned the test booklets along with class lists giving each student's WASL scores. However, for some schools, and for some individual students, this information was not available. For data analysis purposes, students were included in the analyses if they had scores for the analysis in question.

Data Analyses

Data analyses included item analyses, reliability analyses, and validity analyses. Item analyses were conducted before total scores were obtained for strands and totals.

Item analyses

Several types of data were used to evaluate quality of items: item difficulty indices (percent earning each score point), item means, item-test correlations, and correlations between items and criterion tests (WASL test scores for Reading, Writing, and Mathematics).

Item difficulty. Item means were examined to ensure item difficulty was appropriate for the students. Typically, for criterion-referenced assessments that measure students' achievement of grade level appropriate standards, items should be medium to easy in difficulty. If students have learned the targeted knowledge and skills, the items will appear to be easy. Item means should be between half and all points possible for a given item. Another measure of item difficulty is the percent of students earning each score on an item. For an easy four-point item, most of the students will earn scores of 3 and 4. For a difficult four-point item, most of the students will earn scores of 0, 1 and 2.

Item to Total Correlations. Correlations between item scores and the task total indicate whether the item performance is related to the performance on the task as a whole. When correlations are positive and greater than .25, there is a good relationship between performance

on the item and performance on the task. When correlations are close to 0.0, there is little relationship between performance on the item and performance on the task or test. When correlations are negative, students who performed well on the task or test performed poorly on the item and vice versa.

Three types of correlations were obtained in order to evaluate the quality of items. Item scores were correlated with the total scores for the PEI strands, the total EE scores, and with WASL scores. Correlations with PEI strand scores and EE total scores show are measures of how well items fit within the strand or test as a whole. Correlations with WASL scores are measures of the validity of scores. Items should correlate well with WASL tests measuring similar knowledge or skills. Items should correlate poorly with tests measuring very different knowledge or skills. It was expected that the PEI WASL-like items would correlate well with WASL Reading and Mathematics scores and would vary in their relationships with WASL Writing scores depending on the demands of the items. PEI WASL-like scores were expected to have their lowest correlations with WASL Writing scores.

Evidence for Reliability

To obtain evidence for the reliability of the PEI WASL-like scores, two types of data were analyzed: inter-rater agreement data (to look at reliability of raters using rubrics) and internal consistency.

Inter-rater reliability. Measures of inter-rater agreement included: 1) the percent of times two raters gave exactly the same score to students' responses; 2) the percent of times two raters gave the same or adjacent scores to students' responses (for example, Rater 1 might give a student a score of 1 on a response and Rater 2 might give the same response a score of 2), and 3) the correlation between the total scores (sum of the item scores) given by two raters.

Internal consistency. Internal consistency was determined once the final score was assigned for each student's response to each item. Internal consistency is a measure of how consistent students are in their responses across items in a task or test form. This is an index of the reliability of student level scores.

Evidence for validity

To obtain evidence for the validity of the PEI WASL-like assessment scores, the correlations between scores from the PEI WASL-like assessments, the Iowa Test of Basic Skills (ITBS), and WASL were to be examined. However, lack of ITBS scores prevented analysis of this data. WASL-like task scores and PEI total scores were correlated with WASL test scores for Reading, Listening, Mathematics, and Writing.

Environmental education integrates subject matter across traditional discipline boundaries. The items within each PEI WASL-like task assess multiple subject areas. Therefore, all of the items were classified in terms of the knowledge and/or skills measured. Item validity evidence was obtained by correlating each item score with WASL Reading, Mathematics, and Writing Scores. These correlations provide evidence about the relationship between what is measured on the WASL tests and what is measured on the WASL-Like PEI items.

Item scores within each task were summed to create task total scores. The task scores (or benchmark strand scores) from the PEI WASL-like tests were correlated with the scores from the

WASL tests. It was expected that the strand scores would correlate moderately well with WASL tests that required similar knowledge and skills.

A third type of validity evidence was the correlation between PEI total scores and WASL Reading, Listening, Mathematics, and Writing scores. It was expected that, given the multiple content areas represented in the PEI WASL-like assessments, correlations between PEI totals scores and WASL scores would be moderate to moderately high for Reading, Mathematics, and Writing but low for Listening. As with multidimensional items, PEI total scores include multiple dimensions of achievement.

Results

Item Analyses

Item Difficulty

Tables 2 through 12 give the item means and the percent earning each score point. As can be seen, items differed in difficulty across strands and grade levels. Two-point items with means greater than 1.0 would be considered moderately easy. Four-point items with means greater than 2.0 would be considered moderately easy. One-point (multiple-choice) items with means greater than .50 would be considered moderately easy. In tables 2 through 12, moderately easy items are given in bold face.

Table 2

Item Means and Percent Earning Each Score for WASL-like Items: Grade 5, Form A, Inquiry Task

Item Mean	Points Possible	Percent 0	Percent 1	Percent 2	Percent 3	Percent 4
1.29	4	24%	38%	29%	4%	5%

Table 3

Item Means and Percent Earning Each Score for WASL-like Items: Grade 5, Form A, Systems Task

Item Number	Item Mean	Points Possible	Percent 0	Percent 1	Percent 2	Percent 3	Percent 4
1	**1.22**	2	14%	51%	36%		
2	**2.53**	4	2%	21%	20%	35%	22%
3	**1.06**	2	15%	64%	21%		
4	.95	2	34%	37%	29%		
5	.88	2	46%	21%	33%		

Table 4

Item Means and Percent Earning Each Score for WASL-like Items: Grade 5, Form B, Inquiry Task

Item Number	Item Mean	Points Possible	Percent 0	Percent 1	Percent 2	Percent 3	Percent 4
1	**3.45**	4	5%	1%	6%	17%	70%
2	**.68**	1	33%	67%			
3	.25	1	75%	25%			
4	.50	2	64%	21%	15%		
5	**1.01**	2	26%	46%	28%		

Table 5

Item Means and Percent Earning Each Score for WASL-like Items: Grade 5, Form B, Systems Task

Item Number	Item Mean	Points Possible	Percent 0	Percent 1	Percent 2
1	**1.07**	2	19%	55%	26%
2	**1.12**	2	34%	22%	44%
3	.63	2	54%	30%	16%
4	**1.34**	2	19%	27%	54%
5	**.80**	1	21%	79%	
6	**.69**	1	31%	69%	
7	.65	2	48%	41%	11%

Table 6

Item Means and Percent Earning Each Score for WASL-like Items: Grade 5, Form B, Civics Task

Item Number	Item Mean	Points Possible	Percent 0	Percent 1	Percent 2	Percent 3	Percent 4
1	.98	2	25%	53%	22%		
2	**1.06**	2	21%	51%	28%		
3	1.84	4	7%	33%	38%	13%	9%
4	.76	2	42%	41%	17%		
5	1.61	4	27%	16%	28%	26%	3%

Table 7

Item Means and Percent Earning Each Score for WASL-like Items: Grade 8, Form A, Inquiry Task

Item Number	Item Mean	Points Possible	Percent 0	Percent 1	Percent 2	Percent 3	Percent 4
1	1.73	4	14%	28%	37%	15%	6%
2	**2.14**	4	22%	10%	23%	22%	22%
3	**.55**	1	48%	52%			
4	**1.38**	2	12%	39%	49%		
5	**1.29**	2	21%	33%	46%		
6	**1.15**	2	24%	41%	34%		

Table 8

Item Means and Percent Earning Each Score for WASL-like Items: Grade 8, Form A, Systems Task

Item Number	Item Mean	Points Possible	Percent 0	Percent 1	Percent 2	Percent 3	Percent 4
1	1.30	4	23%	31%	38%	6%	1%
2	.74	4	60%	19%	13%	5%	22%
3	**2.31**	4	6%	7%	44%	34%	9%
4	1.88	4	12%	21%	37%	25%	4%
5	1.70	4	17%	22%	38%	20%	3%

Table 9

Item Means and Percent Earning Each Score for WASL-like Items: Grade 8, Form A, Civics Task

Item Number	Item Mean	Points Possible	Percent 0	Percent 1	Percent 2	Percent 3	Percent 4
1	**1.11**	2	22%	45%	33%		
2	**1.33**	2	12%	43%	45%		
3	**2.09**	4	17%	16%	24%	27%	16%
4	**1.60**	2	10%	19%	71%		
5a	1.20	4	35%	28%	24%	7%	5%
5b	**1.09**	2	23%	45%	33%		

Table 10

Item Means and Percent Earning Each Score for WASL-like Items: Grade 8, Form B, Inquiry Task

Item Number	Item Mean	Points Possible	Percent 0	Percent 1	Percent 2	Percent 3	Percent 4
1	1.79	4	7%	31%	41%	17%	4%
2	.32	2	74%	20%	6%		

Table 11

Item Means and Percent Earning Each Score for WASL-like Items: Grade 8, Form B, Systems Task

Item Number	Item Mean	Points Possible	Percent 0	Percent 1	Percent 2	Percent 3	Percent 4
1	1.21	4	26%	41%	21%	11%	1%
2	.55	4	68%	18%	8%	6%	1%
3	**2.16**	4	8%	14%	38%	35%	6%
4	1.55	4	26%	21%	32%	15%	6%
5	1.52	4	27%	22%	27%	18%	5%

Table 12

Item Means and Percent Earning Each Score for WASL-like Items: Grade 8, Form B, Civics Task

Item Number	Item Mean	Points Possible	Percent 0	Percent 1	Percent 2	Percent 3	Percent 4
1	**2.90**	4	7%	4%	21%	28%	40%
2	**2.11**	3	10%	14%	31%	45%	
3	1.79	4	19%	20%	32%	20%	9%
4	1.73	4	19%	25%	28%	21%	7%
5	1.65	4	23%	19%	32%	22%	4%

Item Validity: Item to Score Correlations

Tables 13 through 23 give the correlations between item scores and strand scores on the PEI WASL-like assessments as well as scores on WASL tests. These correlations give information about the validity of item scores. The tables also describe the EALR learning target and the cognitive task required for each item in order to assist in evaluating the item correlations. Correlations that were *expected* to be strongest given the cognitive tasks are bold-faced.

Table 13

Correlations between WASL-like Item Scores, PEI Strand Score and WASL Scores: Grade 5, Form A, Inquiry Task

Item	Cognitive Task	Learning Target	Inquiry Strand	WASL Reading	WASL Writing	WASL Listening	WASL Mathematics
1	Reading background information, design an investigation to answer a research question (analysis)	Science inquiry experimental design	1.00	**.484**	**.420**	.367	**.500**

Table 14

Correlations between WASL-like Item Scores, PEI Strand Score and WASL Scores: Grade 5, Form A, Systems Task

Item	Cognitive Task	Learning Target	Systems Strand	WASL Reading	WASL Writing	WASL Listening	WASL Mathematics
1	Reading a picture, prior knowledge	Geography & Science (landforms)	.276	**.284**	.384	.111	.291
2	Recall of information, analysis and representation	Life science (food web)	.319	.132	**.201**	.151	**.207**
3	Reading a picture, analysis, writing, prior knowledge	Economics (natural resources)	.591	**.449**	**.474**	.224	**.438**
4	Reading picture, analysis, prior knowledge, writing	Economics (natural resources)	.482	**.436**	**.405**	.380	.453
5	Reading picture, prior knowledge, writing	Economics (natural resources)	.494	**.396**	**.364**	.154	.232

Table 15

Correlations between WASL-like Item Scores, PEI Strand Score and WASL Scores: Grade 5, Form B, Inquiry Task

Item	Cognitive Task	Learning Target	Inquiry Strand	WASL Reading	WASL Writing	WASL Listening	WASL Mathematics
1	Classify observations - reading	Science-inquiry	.326	**.173**	.158	.065	.163
2	Multiple choice requiring reading	Science inquiry, human-environmental interaction	.081	**.267**	.243	.038	.180
3	Choose an appropriate investigation	Science-inquiry	-.149	-.075	-.180	-.001	-.157
4	Scientific reading, writing, and supporting claims with evidence	Science inquiry	.298	**.248**	**.293**	.126	.128
5	Scientific reading, analysis and prior knowledge	Science inquiry	.405	**.246**	.187	.129	**.179**

Table 16

Correlations between WASL-like Item Scores, PEI Strand Score and WASL Scores: Grade 5, Form B, Systems Task

Item	Cognitive Task	Learning Target	Systems Strand	WASL Reading	WASL Writing	WASL Listening	WASL Mathematics
1	Reading, prior knowledge, analytical - reasoning	Science: role of organisms in an ecosystem	.477	**.227**	.287	.105	**.382**
2	Analytical – draw diagram	Science: food web	.583	.312	.245	.164	**.348**
3	Reading and prior knowledge	Science & Geography: human-environmental interaction	.453	**.216**	.077	.138	.191
4	Analytical – breakdown a set of tasks to show or describe; prior knowledge	Science – reproduction	.491	.269	**.208**	.044	**.373**
5	Reading	Science – habitat	.416	**.295**	.199	.011	.235
6	Problem-solving; prior knowledge	Science – habitat	.126	.030	.115	.103	**.187**
7	Reading, analysis & writing	Science – habitat	.563	**.409**	**.317**	.165	**.386**

Table 17

Correlations between WASL-like Item Scores, PEI Strand Score and WASL Scores: Grade 5, Form B, Civics Task

Item	Cognitive Task	Learning Target	Civics Strand	WASL Reading	WASL Writing	WASL Listening	WASL Mathematics
1	Reading and prior knowledge	Reading: main ideas and details	.470	**.293**	.217	.362	.220
2	Reading and prior knowledge	Reading: main ideas and details	.550	**.176**	.077	.293	.209
3	Reading and prior knowledge	Reading: main ideas and details	.638	**.342**	.272	.173	.268
4	Computation and showing mathematical procedures	Mathematics: number sense	.228	.281	**.310**	.051	**.343**
5	Reading, organizing and presenting information	Science – human-environmental interaction	.551	**.350**	**.291**	.167	.370

Table 18

Correlations between WASL-like Item Scores, PEI Strand Score and WASL Scores: Grade 8, Form A, Inquiry Task

Item	Cognitive Task	Learning Target	Inquiry Strand	WASL Reading	WASL Writing	WASL Listening	WASL Mathematics
1	Summarizing statistical info	Mathematics: statistics and probability	.453	.066	.020	-.055	**.036**
2	Organizing info, representing info through graphs	Mathematics: communicate through graphs	.476	**.254**	.209	.055	**.300**
3	Analysis of mathematical information	Reading graphs	.368	.129	.099	-.073	**.220**
4	Reading charts, math, prior knowledge	Mathematics: statistics; Science: human-environmental interaction	.476	**.310**	.112	.161	**.220**
5	Analytical reasoning & prior knowledge	Science inquiry: field investigations	.588	.307	.243	-.011	**.416**
6	Prior knowledge	Geography: human environmental interaction; Science human-environmental interaction	.527	.378	.156	.066	.347

Table 19

Correlations between WASL-like Item Scores, PEI Strand Score and WASL Scores: Grade 8, Form A, Systems Task

Item	Cognitive Task	Learning Target	Systems Strand	WASL Reading	WASL Writing	WASL Listening	WASL Mathematics
1	Reading the picture, prior knowledge, analysis of situation	Science: system processes	.559	**.209**	.218	.090	**.327**
2	Prior knowledge; representing information	Science: system processes	.473	**.208**	.114	.099	.167
3	Reading the picture; prior knowledge	Science: habitat	.631	**.155**	.171	.001	.092
4	Reading and analyzing pictures; prior knowledge	Reading: compare & contrast; Mathematics: reasoning; Science: human-environmental interaction	.706	**.199**	.100	-.012	**.273**
5	Reading the picture, analysis, listing, prior knowledge	Science: human-environmental interaction	.677	**.190**	-.014	.023	**.143**

Table 20

Correlations between WASL-like Item Scores, PEI Strand Score and WASL Scores: Grade 8, Form A, Civics Task

Item	Cognitive Task	Learning Target	Civics Strand	WASL Reading	WASL Writing	WASL Listening	WASL Mathematics
1	Reading and analyzing text; prior knowledge	Reading: main ideas & details; Civics: rights & responsibilities of citizens	.478	**.160**	.124	-.040	**.277**
2	Reading and analyzing text; prior knowledge	Reading: main ideas & details; Civics: rights & responsibilities of citizens	.544	**.219**	.174	-.027	**.311**
3	Reading and analyzing text; prior knowledge	Reading: main ideas & details; Civics: constitutional rights; Science: human-environmental interaction	.588	**.298**	.106	.031	**.349**
4	Reading and analyzing text; sorting information ; prior knowledge	Reading: analyzing text; Civics: rights & responsibilities of citizens; Math: sorting & classifying	.382	**.396**	.160	.130	**.445**
5a	Persuasive writing; using evidence to support claims	Writing: persuasive writing; Science: habitat; Social Studies skills: claims & evidence	.562	**.311**	**.234**	.100	.393
5b	Application of skills	Writing conventions	.528	.378	**.197**	.199	.358

Table 21

Correlations between WASL-like Item Scores, PEI Strand Score and WASL Scores: Grade 8, Form B, Inquiry Task

Item	Cognitive Task	Learning Target	Inquiry Strand	WASL Reading	WASL Writing	WASL Listening	WASL Mathematics
1	Design and write steps for a scientific investigation; read and analyze background information	Scientific inquiry	.347	**.297**	**.264**	.174	**.388**
2	Prior knowledge	Science: energy transfer	.347	.197	.102	.198	.267

Table 22

Correlations between WASL-like Item Scores, PEI Strand Score and WASL Scores: Grade 8, Form B, Systems Task

Item	Cognitive Task	Learning Target	Systems Strand	WASL Reading	WASL Writing	WASL Listening	WASL Mathematics
1	Reading the picture, prior knowledge, analysis	Science: system processes	.545	**.287**	.096	.212	**.330**
2	Prior knowledge; representing information, analysis	Science: system processes	.431	**.137**	**.175**	.175	**.280**
3	Reading the picture; prior knowledge	Science: habitat	.590	**.321**	.170	.269	.333
4	Analyzing pictures; prior knowledge	Reading: compare & contrast; Mathematics: reasoning; Science: human-environmental interaction	.642	**.346**	.330	.200	.301
5	Reading the picture, analysis, listing, prior knowledge	Science: human-environmental interaction	.661	**.371**	.318	.225	**.349**

Table 23

Correlations between WASL-like Item Scores, PEI Strand Score and WASL Scores: Grade 8, Form B, Civics Task

Item	Cognitive Task	Learning Target	Civics Strand	WASL Reading	WASL Writing	WASL Listening	WASL Mathematics
1	Analysis of text	Reading: main ideas & details	.620	**.236**	.216	.242	**.207**
2	Analysis of text; prior knowledge	Reading: main ideas & details; Civics: civic action	.733	**.203**	.296	.218	**.212**
3	Analysis of picture	Reading pictures & text	.746	**.321**	.324	.229	**.274**
4	Representing information	Arts: using text and images to communicate ideas responsibly	.755	.293	**.278**	.214	.310
5	Read background information, communicate ideas and plan solutions to problems	Writing: organizing information; Science: design a solution to a problem	.655	**.265**	**.246**	.157	.227

Item Reliability: Inter-rater reliability

Tables 24 through 27 give the rater agreement for items across pairs of raters for the WASL-like items. Rater agreement, also called inter-rater reliability, indicates the percent of times a pair of raters give exactly the same score to the same students' responses. Rater agreement indicates how reliably a scoring rubric is applied to students' responses. Inter-rater reliability data that are of concern are bold-faced.

Some of the flagged items show inter-rater agreement concerns for one pair of raters or for any pairing that includes an individual rater. This may suggest that a particular rater was not sufficiently precise or inadequately trained in applying the rubric to students' work. If an item was flagged for two or three pairs of raters, PEI project coordinators looked closely at the item to determine where the problem lay. As can be seen in Tables 24 through 27, very few items were flagged.

In scoring each item, three raters scored students' work. The final score assigned to a response was based on agreement between at least two raters. A fourth rater, an expert with the PEI WASL-like items and rubrics, resolved discrepancies when two raters did not give the same score. For the item difficulty and item validity analyses, the final item score was used.

Table 24

Rater (R) Agreement by Pair of Raters for Grade 5 Form A WASL-like Items

Strand	Item	Points Possible	R1 & R2 Exact Agreement	R1 & R2 Exact + Adjacent Agreement	R1 & R3 Exact Agreement	R1 & R3 Exact + Adjacent Agreement	R2 & R3 Exact Agreement	R2 & R3 Exact + Adjacent Agreement
Inquiry	1	4	53.6%	93%	61.6%	93%	56.5%	96%
Systems	1	2	95.4%	~100%	96.2%	100%	94.7%	100%
	2	4	72.8%	99%	79.6%	97%	78.8%	97%
	3	2	86.4%	100%	89.3%	~100%	87.8%	~100%
	4	2	90.9%	100%	89.4%	100%	92.5%	~100%
	5	2	91.6%	~100%	94.7%	~100%	92.5%	~100%

Table 25

Exact Agreement by Rater (R) Pair for Grade 5 Form B WASL-like Items

Strand	Item	Points Possible	R1 & R2 Exact Agreement	R1 & R2 Exact + Adjacent Agreement	R1 & R3 Exact Agreement	R1 & R3 Exact + Adjacent Agreement	R2 & R3 Exact Agreement	R2 & R3 Exact + Adjacent Agreement
Inquiry	1	4	77.7%	95%	82.8%	93%	73.6%	93%
	2	1	98.9%	100%	98.9%	100%	100%	100%
	3	1	99.4%	100%	98.9%	100%	98.3%	100%
	4	2	65.4%	94%	48.5%	91%	**55.0%**	**88%**
	5	2	54.4%	97%	43.8%	91%	**36.6%**	**87%**
Systems	1	2	43.6%	93%	48.8%	94%	59.7%	96%
	2	2	58.1%	90%	66.7%	93%	60.6%	96%
	3	2	56.3%	87%	45.3%	94%	55.5%	97%
	4	2	**29.4%**	**77%**	**35.0%**	**79%**	66.8%	99%
	5	1	98.3%	100%	**76.9%**	**81%**	**77.7%**	**81%**
	6	1	98.3%	100%	96.6%	100%	96.6%	100%
	7	2	52.1%	98%	59.9%	94%	58.1%	99%
Civics	1	2	83.6%	100%	79.7%	99%	85.3%	100%
	2	2	80.4%	100%	82.0%	100%	83.2%	100%
	3	4	78%	98%	62.8%	96%	70.5%	96%
	4	2	92.8%	99%	94.0%	99%	97.7%	100%
	5	4	64.4%	96%	51.7%	89%	68.9%	97%

Table 26

Exact Agreement by Rater (R) Pair for Grade 8 Form B WASL-like Items

Strand	Item	Points Possible	R1 & R2 Exact Agreement	R1 & R2 Exact + Adjacent Agreement	R1 & R3 Exact Agreement	R1 & R3 Exact + Adjacent Agreement	R2 & R3 Exact Agreement	R2 & R3 Exact + Adjacent Agreement
Inquiry	1	4	**50.2%**	**85%**	53.1%	92%	**52.7%**	**87%**
	2	4	59.3%	91%	61.2%	92%	61.5%	91%
	3	1	80.6%	94%	87.7%	~100%	86.9%	~100%
	4	2	60.2%	94%	64.6%	98%	62.9%	97%
	5	2	70.7%	96%	54.9%	93%	60.8%	99%
	6	2	67.5%	98%	69.3%	98.5%	70.1%	99%
Systems	1	4	68.5%	98%	81.9%	100%	69.1%	99%
	2	4	62.3%	95%	77.2%	94%	62.4%	93%
	3	4	70.1%	96%	70.1%	96%	62.4%	98%
	4	4	63.4%	96%	81.4%	96%	62.0%	98%
	5	4	62.9%	97%	80%	97%	62.9%	98%
Civics	1	2	60.2%	99%	63.8%	96%	66.3%	97%
	2	2	68.4%	~100%	73.6%	~100%	62.7%	99%
	3	4	47.7%	89%	52.4%	92%	53.0%	96%
	4	2	88.6%	100%	88.0%	100%	90.1%	100%
	5	4	61.7%	93%	62.7%	94%	63.3%	94%
	6	2	72.5%	99%	66.8%	97%	68.9%	97%

Table 27

Exact Agreement by Rater (R) Pair for Grade 8 Form B WASL-like Items

Strand	Item	Points Possible	R1 & R2 Exact Agreement	R1 & R2 Exact + Adjacent Agreement	R1 & R3 Exact Agreement	R1 & R3 Exact + Adjacent Agreement	R2 & R3 Exact Agreement	R2 & R3 Exact + Adjacent Agreement
Inquiry	1	4	78.9%	98%	79.6%	99%	84.8%	~100%
	2	2	84.4%	~100%	86.9%	~100%	90.7%	100%
Systems	1	4	72.5%	99%	70%	98%	75.3%	100%
	2	4	75.8%	98%	83.8%	97%	78.2%	97%
	3	4	68.5%	99%	78.1%	99%	69.2%	96%
	4	4	**60.7%**	**89%**	78.9%	97%	70.8%	90%
	5	4	61.7%	95%	77.4%	98%	64.7%	95%
Civics	1	4	75.7%	99%	76.3%	~100%	70.1%	~100%
	2	3	70.2%	~100%	74.7%	99%	74.4%	~100%
	3	4	66.7%	95%	66.9%	97%	63.5%	96%
	4	4	66.2%	97%	68.5%	95.8%	70.3%	98%
	5	4	77.4%	97%	65.2%	94%	69.2%	98%

Another measure of inter-rater reliability is the degree to which students would earn the same total score if scored by different raters. Although there may be differences among raters, particularly for students whose responses are on a borderline between two score points on a scoring rubric, 'error' should be randomly positive and negative. Therefore, raters should randomly assign borderline papers the higher or lower scores. Tables 28 through 38 give the correlations between pairs of raters at the total scores level for each PEI WASL-like strand.

Table 28

Correlation between Pairs of Raters (R) at the Total Score Level, Grade 5, Form A Inquiry Strand*

	Rater 1	Rater 2	Rater 3
Rater1	1.00	.721	.770
Rater 2		1.00	.770
Rater 3			1.00

* There was one item in the Inquiry Strand for Grade 5, Form A

Table 29

Correlation between Pairs of Raters (R) at the Total Score Level, Grade 5, Form A Systems Strand

	Rater 1	Rater 2	Rater 3
Rater1	1.00	.960	.952
Rater 2		1.00	.955
Rater 3			1.00

Table 30

Correlation between Pairs of Raters (R) at the Total Score Level, Grade 5, Form B Inquiry Strand

	Rater 1	Rater 2	Rater 3
Rater1	1.00	.714	.713
Rater 2		1.00	.638
Rater 3			1.00

Table 31

Correlation between Pairs of Raters (R) at the Total Score Level, Grade 5, Form B Systems Strand

	Rater 1	Rater 2	Rater 3
Rater1	1.00	.736	.742
Rater 2		1.00	.774
Rater 3			1.00

Table 32

Correlation between Pairs of Raters (R) at the Total Score Level, Grade 5, Form B Civics Strand

	Rater 1	Rater 2	Rater 3
Rater1	1.00	.928	.883
Rater 2		1.00	.933
Rater 3			1.00

Table 33

Correlation between Pairs of Raters (R) at the Total Score Level, Grade 8, Form A Inquiry Strand

	Rater 1	Rater 2	Rater 3
Rater1	1.00	.752	.826
Rater 2		1.00	..812
Rater 3			1.00

Table 34

Correlation between Pairs of Raters (R) at the Total Score Level, Grade 8, Form A Systems Strand

	Rater 1	Rater 2	Rater 3
Rater1	1.00	.902	.941
Rater 2		1.00	.870
Rater 3			1.00

Table 35

Correlation between Pairs of Raters (R) at the Total Score Level, Grade 8, Form A Civics Strand

	Rater 1	Rater 2	Rater 3
Rater1	1.00	.852	.888
Rater 2		1.00	.866
Rater 3			1.00

Table 36

Correlation between Pairs of Raters (R) at the Total Score Level, Grade 8, Form B Inquiry Strand*

	Rater 1	Rater 2	Rater 3
Rater1	1.00	.868	.879
Rater 2		1.00	.923
Rater 3			1.00

* There were two items in the Inquiry Strand for Grade 8, Form B

Table 37

Correlation between Pairs of Raters (R) at the Total Score Level, Grade 8, Form B Systems Strand

	Rater 1	Rater 2	Rater 3
Rater1	1.00	.918	.946
Rater 2		1.00	.927
Rater 3			1.00

Table 38

Correlation between Pairs of Raters (R) at the Total Score Level, Grade 8, Form B Civics Strand

	Rater 1	Rater 2	Rater 3
Rater1	1.00	.954	.966
Rater 2		1.00	.958
Rater 3			1.00

5th Grade Item Analysis

The item analyses for the Grade 5 WASL-like items are given in Tables 2 through 6, Tables 13 through 17, Tables 24 and 25, 28 through 32. Tables 2 through 6 show that, while some items appeared to be difficult, many item means were within the criterion of half to all possible points. Of particular difficulty were items requiring prior knowledge, an item requiring students to design an investigation, and items that required students to make claims and support them with evidence. These results are consistent with that found on WASL science assessments.

Correlations between Grade 5 WASL-like items and EE strand scores and WASL test scores (Tables 13 through 17) helped to identify items that did not relate to the overall strand or the relevant subject area. Form B, Inquiry items 2 and 3 and Systems item 6 were poorly related to the total strand score. All three were multiple-choice items and were eliminated from the strand score totals when examining total score validity results.

It was expected that the strongest correlations would be between item scores and scores for WASL Reading, Mathematics, and Writing – depending on the demands of items. Correlations with WASL Listening scores were expected to be low since the WASL-like items do not require listening – lending evidence for the validity of the PEI WASL-like items. Correlations were examined to determine whether items correlated in expected ways. Items with poor item score to strand score correlations were removed from the total strand scores prior to obtaining correlations between PEI strands scores and WASL test scores.

With few exceptions, items requiring analysis skills – regardless of the focus of the item – tended to correlate best with WASL Mathematics scores; items requiring reading skills correlated well with WASL Reading scores; items requiring writing or organizing and representing information correlated well with WASL Writing scores. Correlations between item scores and WASL Listening scores were, with few exceptions, the lowest correlations across all items.

The inter-rater reliability (Tables 24 and 25) information identified several items that had low inter-rater agreement. These items tended to be 4 point items. As mentioned earlier, rater

agreement is difficult when student responses are on a borderline between two score levels. Therefore, when there are more points for an item, more borderline responses are possible. It is important to look, not only at the exact agreement, but the exact and adjacent agreement percents. In most cases, exact plus adjacent rater agreements for all Grade 5 items were greater than 90% suggesting that rater agreement was acceptable.

Rater agreement at the total score level (Tables 27 through 32) suggested that the Inquiry strand score for Grade 5 Form A was not very reliable. There was only one item in the strand score. In such cases, random differences on borderline responses cannot be canceled out with scores on other items. This suggests that additional items should be added to the Inquiry task. Total score rater agreement was also low for the Inquiry and Systems scores for Grade 5 Form B. Recall that these are the two strands that had faulty multiple choice items; therefore, the rater agreement is likely to be higher once these faulty items are removed from the analyses.

8th Grade Item Analysis

The item analyses for the Grade 8 WASL-like items are given in Tables 7 through 12, Tables 18 through 23, Tables 26 and 27, 33 through 38. Tables 7 through 12 show that student responses to many items met the criterion of at least 50% of the possible points. Items that posed the most difficulty were those requiring prior knowledge. For example, items 1 and 2 in Grade 8, Forms A and B required students to identify land forms in a picture and describe the natural process that brought them about. Item 5a in Grade 8 Form A required students to identify what could be done to prevent the negative impacts of human behaviors – again requiring prior knowledge. These results are consistent with that found on WASL science assessments.

Correlations between Grade 8 WASL-like items and EE strand scores and WASL test scores (Tables 18 through 23) helped to identify items that did not relate to the overall strand or the relevant subject area. Form A, Inquiry item 1 had a moderate correlation with the strand score. Students earned 1.73 out of 4 points on the item. Students earned all possible score points. It is possible that this item was less discriminating because summary of data is a fairly straightforward task; therefore, students at all skill levels were able to complete the item. There were no clearly flawed Grade 8 items based on the item to strand total correlations; however, Item 1 from the Grade 8 Form A Inquiry strand was deleted from the overall strand score.

As with fifth grade, it was expected that the strongest correlations would be between Grade 8 item scores and WASL scores for Reading, Mathematics, and Writing – depending on the demands of items. Correlations with WASL Listening scores were expected to be low since the WASL-like items do not require listening – lending evidence for the validity of the PEI WASL-like items. Correlations were examined to determine whether items correlated in expected ways. Items with poor item score to strand score correlations were removed from the total strand scores prior to obtaining correlations between PEI strands scores and WASL test scores.

With few exceptions, items requiring analysis and mathematical skills – regardless of the focus of the item – tended to correlate well with WASL Mathematics scores; items requiring reading skills correlated well with WASL Reading scores; items requiring writing or organizing and representing information correlated well with WASL Writing scores. Correlations between item scores and WASL Listening scores were, with few exceptions, the lowest correlations across all items.

The inter-rater reliability information (Tables 26 and 27) identified several items that had low inter-rater agreement. As with grade 5, these items tended to be 4 point items since raters are less likely to agree on responses that are on a borderline between two score levels. In most cases, exact plus adjacent rater agreements for all Grade 8 items were greater than 90% suggesting that rater agreement was acceptable. The item with lowest exact plus adjacent agreement was Item 1 in the Inquiry strand for Grade 8 Form A. This item was removed from the strand score for poor validity correlations.

Rater agreement at the total score level (Tables 33 through 38) suggested that the Inquiry strand score for Grade 8, Form A was not very reliable. This again reflects the statistical flaws in Item 1 of the Grade 8 Form A Inquiry strand. All other rater agreement correlations at the total score level were acceptable given the limited number of items in each strand.

Evidence for the Reliability of PEI Strand and Total Scores

Score reliability for PEI strand and PEI total scores were obtained through the use of the Alpha Coefficient, a measure of internal consistency in student responses. Tables 39 through 42 present the means, standard deviations, and alpha coefficients for each PEI strand score and the PEI total score for each test form.

Table 39

Means, Standard Deviations, and Alpha Coefficients for each Strand Score and the PEI Total Score, Grade 5, Form A

Score	Mean	Standard Deviation	Alpha Coefficient
Inquiry Task	1.29	1.04	NA
Systems Task	6.64	2.70	.67
PEI Total	8.10	3.33	.73

Table 40

Means, Standard Deviations, and Alpha Coefficients for each Strand Score and the PEI Total Score, Grade 5, Form B

Score	Mean	Standard Deviation	Alpha Coefficient
Inquiry Task	4.98	1.86	.57
Systems Task	5.62	2.85	.76
Civics Total	6.26	3.07	.72
PEI Total	16.81	5.58	.78

Table 41

Means, Standard Deviations, and Alpha Coefficients for each Strand Score and the PEI Total Score, Grade 8, Form A

Score	Mean	Standard Deviation	Alpha Coefficient
Inquiry Task	6.48	2.94	.74
Systems Task	7.93	3.88	.82
Civics Total	8.42	3.69	.75
PEI Total	22.85	8.49	.87

Table 42

Means, Standard Deviations, and Alpha Coefficients for each Strand Score and the PEI Total Score, Grade 8, Form B

Score	Mean	Standard Deviation	Alpha Coefficient
Inquiry Task	2.11	1.27	.52
Systems Task	6.98	3.99	.79
Civics Total	10.17	4.70	.87
PEI Total	19.51	8.92	.90

The alpha coefficients range from .52 for the Grade 8, Form B Inquiry strand score (composed of 2 items) to .87 for the Grade 8, Form B Civics strand total score (composed of 5 items). Most alpha coefficients for PEI strand scores are in the mid .70s. This suggests that most strand scores are not reliable enough to stand alone. However, the PEI total score alpha coefficients are more acceptable (.73 to .90) given the number of items on each test forms. The lowest alpha coefficient is for Grade 5, Form A which was composed of only two strands: Inquiry and Systems. The results across the four test forms suggest that students completing these test forms would be likely to obtain similar scores if they completed the forms again.

Evidence for the Validity of Strand and Total Scores

Correlations between PEI strand and total scores and WASL total scores were examined to determine whether the scores correlated expected ways. Tables 43 through 46 present the correlations among strand and total scores for each of the PEI test forms.

Correlations among scores provide several points of information regarding the validity of scores for the PEI WASL-like assessments. First, correlations among the different strands within each form are moderate (.312 to .670) with most in the 50s and 60s. This suggests that each strand taps into a different body of knowledge and skills. Inquiry items generally require science knowledge and skills, Civics items generally require social science knowledge and skills; Systems items generally require both science and social science knowledge and skills.

Scores between PEI WASL-like strand scores and WASL scores are informative. For example, the correlation between the Grade 5, Form A Inquiry strand scores and WASL Mathematics scores are stronger than correlations between the Inquiry strand scores and the other WASL scores. The strand consists of one item (Let the Sunshine In! – Item1) that requires analysis and design of a scientific investigation. For three out of four of the forms, the Inquiry strand correlates best with WASL Mathematics. In the third case (Grade 5, Form B), the Inquiry task requires students to examine written observations made by other students; therefore, scores for this task correlate better with WASL Reading scores.

Table 43

Correlations between Strand Scores, PEI Total Scores, and WASL Test Scores, Grade 5, Form A

Score	Inquiry Strand	Systems Strand	PEI Total	WASL Mathematics	WASL Reading	WASL Writing	WASL Listening
Inquiry Strand	1.00	.469	.694	.500	.484	.420	.367
Systems Strand		1.00	.961	.459	.477	.513	.293
PEI Total			1.00	.542	.563	.568	.357

Table 44

Correlations between Strand Scores, PEI Total Scores, and WASL Test Scores, Grade 5, Form B

Score	Inquiry Strand	Systems Strand	Civics Strand	PEI Total	WASL Mathematics	WASL Reading	WASL Writing	WASL Listening
Inquiry Strand	1.00	.469	.312	.644	.216	.299	.281	.154
Systems Strand		1.00	.399	.823	.488	.423	.328	.160
Civics Strand			1.00	.787	.424	.435	.350	.299
PEI Total				1.00	.533	.550	.444	.291

Table 45

Correlations between Strand Scores, PEI Total Scores, and WASL Test Scores, Grade 8, Form A

Score	Inquiry Strand	Systems Strand	Civics Strand	PEI Total	WASL Mathematics	WASL Reading	WASL Writing	WASL Listening
Inquiry Strand	1.00	.558	.604	.832	.395	.340	.230	.003
Systems Strand		1.00	.470	.827	.289	.274	.171	.057
Civics Strand			1.00	.850	.477	.383	.212	.082
PEI Total				1.00	.459	.303	.263	-.003

Table 46

Correlations between Strand Scores, PEI Total Scores, and WASL Test Scores, Grade 8, Form B

Score	Inquiry Strand	Systems Strand	Civics Strand	PEI Total	WASL Mathematics	WASL Reading	WASL Writing	WASL Listening
Inquiry Strand	1.00	.497	.576	.693	.415	.314	.242	.223
Systems Strand		1.00	.670	.891	.427	.398	.291	.289
Civics Strand			1.00	.925	.301	.323	.321	.259
PEI Total				1.00	.404	.400	.336	.294

Systems strand scores generally correlate best with WASL Reading and Mathematics scores. In responding to Systems items, students must analyze information presented pictorially or in text; therefore, reading is a requisite skill. In addition, Systems tasks require students to use diagrams to demonstrate cycles within systems. The analytical skills required may be similar to those required for mathematics items. For only one task (Grade 5, Form A) did the WASL writing scores correlate strongly with the Systems strand scores (.513). This task required students to represent their ideas both in writing and in diagrams. Therefore, both the WASL Writing assessment and the Grade 5, Form A, PEI Systems task require students to organize and represent information.

PEI Civics strand scores also correlate best with WASL Reading, Writing, and Mathematics scores. The Grade 5, Form B Civics items require reading comprehension and computation; therefore, correlations with reading and mathematics make sense. The fairly strong correlation between scores for the Grade 8, Form A Civics strand and WASL Mathematics scores is more difficult to understand. The task requires reading and analysis of text as well as social studies prior knowledge. Finally, the correlations between the Grade 8, Form B Civics strand scores and WASL Reading and Writing make sense given that the students must read and analyze text and then write a letter in which they recommend environmentally responsible behaviors.

Correlations between PEI strand scores and WASL Listening scores are uniformly low which was expected given that the PEI items did not require listening comprehension skills. In general, the correlations among WASL scores and the PEI WASL-like strand scores were in expected directions and of expected magnitudes for the different WASL content areas. These relationships provide good evidence for validity of scores from PEI assessments as measures of the similar skills as the WASL.

Discussion

This research investigated the validity and the reliability of the PEI WASL-like assessments. Analyses included item analyses to select the best items from each task, inter-rater reliability analyses, item validity analysis, strand and total score reliability analyses and strand score validity analyses.

In general, the short-answer and extended response items in the WASL-like tasks functioned well. Items requiring prior knowledge were the most difficult even though the content assessed was aligned with the state science and social studies curriculum standards. Items requiring reading of text or pictures and written analyses of given information were easier. Items that required design of a scientific investigation were very difficult; however, these data are consistent with that found on Washington State's WASL Science test.

Inter-rater reliability data was at acceptable levels for all items and tasks, except for 4 items that were removed from further analyses. Item validity analyses provided support for the validity of item scores. Correlations between item scores and WASL scores generally made sense in terms of the types of thinking and skills required to complete the items.

Alpha coefficients for strand scores were low enough, especially for tasks with only two or three items, that use of these tasks alone is not warranted. Alpha coefficients for PEI total scores were at acceptable levels. The desired alpha coefficient is greater than .80. Neither of the total scores for the Grade 5 forms reached this level. Grade 5, Form A was short one task; therefore, it was not a complete test. Grade 5, Form B was shorter than the Grade 8 forms due to

the loss of 3 poor quality items. If the faulty items are replaced with better functioning items, Grade 5, Form B is likely to be as reliable as the Grade 8 forms. Both Grade 8 forms had acceptable levels of internal consistency.

Intercorrelations among PEI strand scores and WASL test scores were moderate to moderately strong for WASL Reading, Mathematics, and Writing scores. The patterns of the correlations generally made sense given the demands of the items in each PEI task. The low correlations with WASL listening scores were expected. The patterns of correlations provide evidence for the validity of the PEI WASL-like assessments as integrated measures. It is expected that correlations with WASL Science scores will be stronger than with the WASL Reading, Mathematics, and Writing scores given the science concepts and skills required by many of the items.

Conclusion

The EE Integrated Benchmarks and associated WASL-like tests are tools that fit with EE instructional programs and Washington State learning standards. They provide K-12 educators, and EE programs with a means to evaluate the ability to improve student achievement in three key areas: 1) valued environmental literacy performances; 2) critical thinking processes; and 3) discipline-based content and skills. This report provides evidence that these assessments are rigorous measures and are of sound technical quality. Effective use of these assessments depends on careful records of students' item scores in order to assess the discrete skills of reading, writing, and mathematics inherent in different items within each integrated task.

In a climate in which schools are being required to test students in so many subject areas, the PEI WASL-like assessments offer the possibility of using a universal context like the environment to integrate discipline standards. Educators will have access to student data in individual subject areas as well as student data on the ability to synthesize and apply information. Over time, as educators implement PEI tools in their programming both in formal classrooms and informal courses, the construct variables embedded in the benchmark performances will be clarified. This clarification leads to developing WASL-like tasks that assess desired performances more effectively.

References

American Forest Foundation. 2003. *Project Learning Tree Pre K – 8 Activity Guide*. Washington D.C.

Council on Environmental Education. (2002). *Project WILD K-12 Curriculum and Activity Guide*. Houston, TX.

Lieberman, G. A, and L. L. Hoody. (1998 & 2002). *Closing the achievement gap: Using the environment as an integrating context for learning*. State Education and Environment Roundtable.

Lieberman, G. A., Hoody, L. and G. M. Lieberman. (2000). *California student assessment project. The effects of environment-based education on student achievement*. State Education and Environment Roundtable.

National Environmental Education and Training Foundation. (2002). *Environmental Learning in America: Working Toward Nationwide Environmental Literacy*. Washington D.C. www.neetf.org

National Research Council. (2001). Knowing what students know the science and design of educational assessment. Committee on the foundations of assessment. Pelligrino, J., Chudowsky, N. and Glaser, R. Eds. Washington D.C.: National Academy Press.

North American Association for Environmental Education (2001) *Using environment-based education to advance learning skills and character development* Washington D.C, National Environmental Education and Training Foundation.

Marzano, R.J., D.J. Pickering, and J.E. Pollock. 2001. *Classroom Instruction that Works: Research-Based Strategies for Increasing Student Achievement*. ASCD. Alexandria, VA.

Office of the Superintendent of Public Instruction (2002). *Washington Assessment of Student Learning, Technical Report for 2000, Grade 4*. Olympia, WA: Author.

Resnick, D.P., & Peterson, N.L. (1991). Evaluating progress toward goal five: A report to the National Center for Education Statistics. Report No. TM 018 020. Washington, DC: Assessing higher order thinking and communication skills in college graduates. ERIC Document Reproduction Service No. ED 340-764

Shepard, L. A. (1989). Why we need better assessments. *Educational Leadership, 46*(7), 4−9.

Smith, M. L. (1991). Meanings of test preparation. *American Educational Research Journal, 28* (3), 521-542.

Sternberg, R. and Spear-Swerling, L. (1996). *Teaching for thinking*. In *Psychology in the classroom: A series on applied educational psychology (*Eds) McCombs, B., and McNeely, S., Washington D.C.: American Psychological Association.

Taylor, C. S. (2002). Incorporating classroom-based assessments into large-scale assessment programs. In T. Haladyna & G. Tindall (Eds.) *Large-scale assessment programs for all students: Development, implementation, and analysis*. Mahwah: NJ: Lawrence Erlbaum Associates, Publishers.

Taylor, C. S. & Nolen, S. B. (2005). *Classroom Assessment: Supporting Teaching and Learning in Real Classrooms*. Columbus, OH: Merrill-Prentice Hall.

The Watercourse and Council for Environmental Education. (1999). *Project WET Curriculum and Activity Guide.* Bozeman, MT.

Volk, T. and McBeth, W. (1998). *Environmental literacy in the United States*. In *Essential readings in environmental education (*Eds) Hungerford, H., Bluhm, W., Volk, T., Ramsey, J. Champaign IL: Stipes Publishing L.L.C.

Wiggins, G., & McTighe, J. (1998) Understanding By Design. Alexandria, VA: Association for Supervision and Curriculum Development.

Appendix A

Example Benchmark Performance Description

Middle School

Civic Participation

Integrated Essential Academic Learning Requirements with the Environment as an Organizing Context

8th grade Benchmark Performance

Civic Participation

By the end of middle school, students will (with guidance from a teacher) work with others to identify a local environmental issue (a problem, or its solutions, for which differing beliefs and values exist, usually involving two or more parties who don't agree. *If students don't understand the beliefs and values of the disagreeing parties, they won't understand the concept of an environmental issue*. See glossary for more information on problem, issue, values) that involves the interplay between human and natural systems. Students will identify stakeholders (those who stand to gain or lose from the resources in the system) and their roles, responsibilities, and perspectives. Students will collaborate to develop a plan to address the issue, create relevant products, and implement the plan. Individually students will prepare a rationale and evaluate the plan's effectiveness. Students will share their work in a chosen format with an appropriate audience.

Performance / Product:

To accomplish this, the student will complete the following steps:

STEP 1 Work with peers and or community members[i] to identify a local environmental issue that involves the interplay between human systems and ecosystems[ii].

1.1 Gather information about the situation from a variety of sources (texts, internet, television, radio, print media, visiting experts)[iii]

1.2 Identify and describe the components of the ecosystem under consideration

1.3 Predict the impacts of the situation on the ecosystem, the economy and human health[iv].

STEP 2 Identify the major players/stakeholders (government agencies, diverse cultural groups, producers consumers, organizations, individuals etc.), their roles, responsibilities and perspectives[v].

2.1 Research two or more stakeholder groups in which they:

a) Examine and describe why the stakeholders hold their perspectives (values and beliefs) by researching the history of these stakeholders and how these histories led to current perspectives.

b) Examine and describe stakeholders' proposals to address the situation

STEP 3 Create, explore, and consider alternative solutions related to the situation.

3.1 Examine others' action plans relevant to the issue.[vi] (e.g. agencies or organization responsible for management or regulation of the ecosystem or human system often have developed management goals)

3.2 Describe ways in which the actions other took impacted the issue.

3.3 Consider whether to modify an existing action plan or develop new plan of action.

STEP 4 Using information from the student's own or their peer's research, develop a plan of action.

4.1 Develop a plan of action including objectives, timelines, tasks, costs, and division of labor.

4.2 Determine what products to create and/or projects to undertake. Plans may include diverse products (proposals for a healthier product[vii], artistic interpretations about the issue[viii], products designed to influence consumers [adults, children, and youth][ix], proposals for strategies to improve the health of the environment[x], proposals for policy change[xi], products designed to influence decision-makers (e.g., politicians, business owners/managers, employees, consumers [adults, children, and youth])[xii], proposal for projects that physically improve the environment.

4.3 Work with others to develop a strategy for maintaining the plan, if needed.

4.4 Develop criteria to evaluate the effectiveness of the plan and products (e.g., impact on the environment, economics, human health; impact on target audience; response of stakeholders; consequences of implementation).

STEP 5 Individually prepare a rationale for the plan[xiii].

5.1 Defend the plan based on facts, data, personal values, and ethical considerations.

5.2 Seek and obtain approval of plan from the people who are in positions to support or help implement the plan.

STEP 6 Implement the plan

6.1 Carry out the details of the plan in collaboration with other participants. Involve local media of appropriate.

6.2 Maintain journal of what happens while implementing the plan

6.3 Collect evidence of civic behaviors including artifacts such as: communications with stakeholders[xiv], videotapes[xv], survey of change of behavior or attitude[xvi], photographs[xvii], products of the plan[xviii], action logs[xix], etc.

STEP 7 Prepare a written[xx] or oral[xxi] report in which they evaluate and summarize the implementation of the plan:

7.1 Describe the environmental situation, the plan to address it, and the rationale for plan

7.2 Summarize the outcome(s) of the plan[xxii]

7.3 Evaluate the plan's effectiveness,[xxiii] based on criteria developed.

7.4 Provide evidence of civic behaviors including artifacts such as: communications with stakeholders[xxiv], videotapes[xxv], survey of change of behavior or attitude, photographs[xxvi], products of the plan[xxvii], action logs[xxviii], etc.

7.5 Select a written or oral format (PowerPoint presentation, written report, poster presentation, etc.)

7.6 Use appropriate conventions of technology, writing and/or speaking/presenting.

8 Use language appropriate to the audience.

Endnotes indicate the specific Essential Academic Learning Requirements (EALRs) that can be demonstrated through this step in the performance.

---

[i] Communication EALRs 3.1.1-4, 3.2.1-3, 3.3.1-4
[ii] History EALR 2.1.1; Science EALRs 1.2.1, 1.3.10
[iii] Reading EALRs 3.1.2-4
[iv] Health/Fitness EALR 3.1.1
[v] Reading EALRs 2.1.2, 2.3.4; History EALRs 2.1.2, 2.3.1; Civics EALR 4.1.1
[vi] Geography EALRs 3.1.1-2, 3. 2.2, 3.3.1; Civics EALR 1.3.2; History EALRs 2.3.1-2
[vii] Science EALR 2.2.2; Health/Fitness EALRs 3.1.1, 3.2.2
[viii] Arts EALRs 3.1.1, 4.1.1
[ix] Communication EALR 2.1.3; Writing EALRs 2.1.2, 2.2.2, 2.2.5, Art EALRs 3.1.1, 4.1.1
[x] Health/Fitness EALRs 3.1.1, 3.2.2; Science EALR 1.3.10
[xi] Civics EALRs 4.2.1, 4.3.1
[xii] Communication EALR 2.1.3; Writing EALRs 2.1.2, 2.2.2, 2.2.5
[xiii] Mathematics EALRs 3.3.2; Civics EALRs 1.2.1, 4.1.1
[xiv] Communication EALR 2.1.1; Writing EALR 2.1.1
[xv] Communication EALRs 2.5.1-2
[xvi] Mathematics EARL 1.4.5
[xvii] Communication EALRs 2.5.1-2
[xviii] Communication EALRs 2.5.1; Arts EALRs 3.1.1, 4.1.1
[xix] Writing EALRs 2.2.4, 2.3.1; History EALRs 2.1.2, 2.2.1
[xx] Writing EALRs 1.1.1-7, 1.2.1-3, 1.2.5-6, 2.2.2, 2.3.1; Science EALRs 2.1.3, 2.1.5; History EALRs 2.3.1-2
[xxi] Communication EALRs 2.1.1-4, 2.2.1-4; Science EALRs 2.1.3, 2.1.5; History EALRs 2.3.1-2
[xxii] Science EALR 2.1.5
[xxiii] Science EALR 2.2.3
[xxiv] Communication EALR 2.1.1
[xxv] Communication EALRs 2.5.1-2
[xxvi] Communication EALRs 2.5.1-2
[xxvii] Communication EALRs 2.5.1; Arts EALRs 3.1.1, 4.1.1
[xxviii] Writing EALRs 2.2.4, 2.3.1; History EALR 2.1.2, 2.2.1

Appendix B

Example WASL-Like Assessment

Grade 8

Civics Strand

Poster Campaign

You want to help people keep the place where we live healthy. Read the article "Washington Weed Whackers" from *Time Magazine for Kids* special issue and decide how you can make a difference in your neighborhood.

WASHINGTON WEED WACKERS

David Bjerklie

An alien species is creeping along the shores of Puget Sound, and this class wants to stop it. Poor Spartina! It's a perfectly good plant where it belongs, but in Washington State's Puget Sound, it has turned into a fast spreading, life-choking weed. A living thing in the wrong environment can often become a pest. That's what a class of fifth- and sixth-graders at Lincoln Elementary School in Mount Vernon, Washington, learned when they began investigating the spartina invasion and its impact on their own part of the Sound, an area called Padilla Bay.

Ecologists call plants and animals that wind up in the wrong place *aliens* or *exotics*. Spartina, or cordgrass, is native to many East Coast waterways, but in Washington State it is an alien species.

AN ALIEN ATTACKS

Spartina spreads easily. In many parts of Puget Sound, it has crowded out native plants like eelgrass, pickleweed, and salt grass. Making matters worse, spartina grows in thick clumps. "It clogs up all the mud and changes the shape of the mud flats," explains student Seth Morris. This affects the entire ecosystem. As native plants and the gentle slopes of the mud flats disappear, native animals like crabs, snails, salmon and shorebirds have less to eat and move out of the area.

When the kids at Lincoln Elementary School decided to take on the spartina problem, they didn't know anything about the plant or how it got to the Northwest. Neither did their teachers Teresa Vaughn and Michael Guelker-Cone. The class hit the books to do some research. Then they contacted local experts. The students searched the Internet and even found a kayak club that was mapping spartina along the shores of Puget Sound.

WHERE DID IT COME FROM?

Seth and Anna Hansen reported on the history of the weed. "Spartina goes back to the late 1800s, when it came here from the East Coast," Seth explains. Settlers who wanted to raise oysters brought the shellfish west packed in wet spartina to keep the oysters fresh. When the oysters were put in new beds in Puget Sound, spartina seeds that had come along for the ride sprouted.

Spartina was also introduced to the area on purpose. Duck hunters planted spartina in hope of attracting more ducks. Engineers brought it in to control erosion, and farmers planted it to feed their cattle.

Some members of the class tackled other aspects of the weed problem. Allison Hamburg, Kyle Brown, Michael Bazan, and Eboni Washington researched Padilla Bay. They drew diagrams of spartina and its effects on the shore and outlined possible solutions. Amy Phillip, Michael Marsh, and Kayla King worked on press releases and publicity. The entire class wrote letters to the state legislature, urging them to control spartina.

Some kids were inspired to write poems about Padilla Bay's natural beauty. One group wrote:

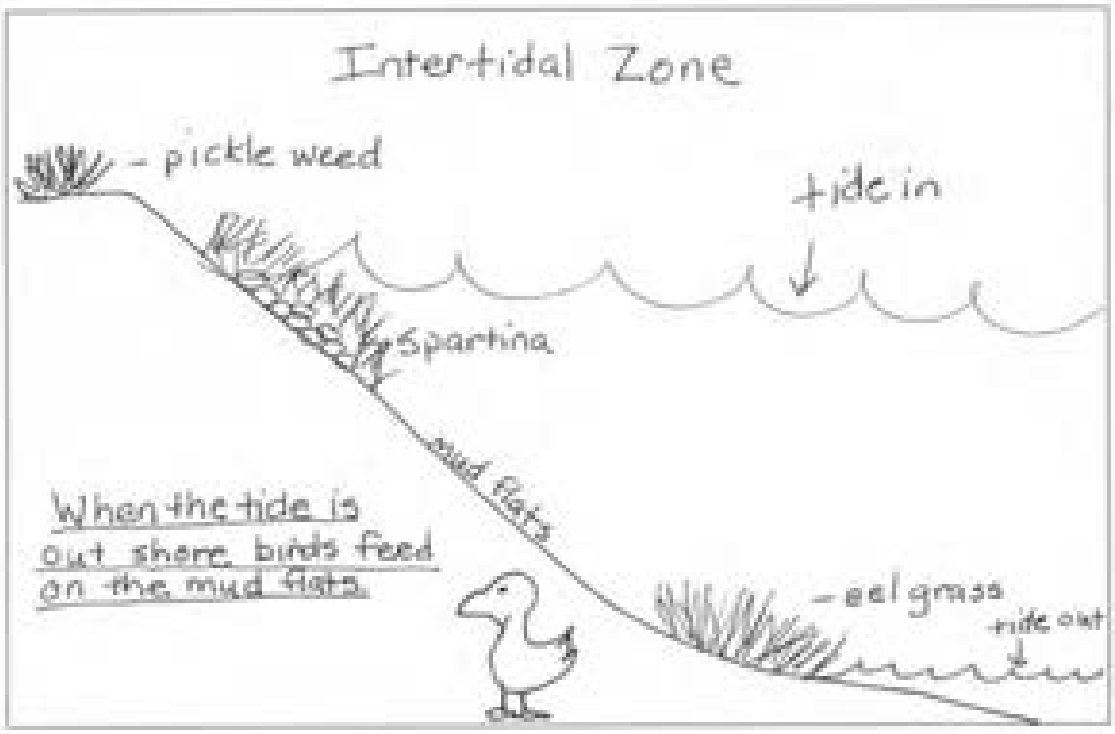
"Elegant fronds of ghostly eelgrass
Gently sway in forests just kissed by the sun
A delicate brew of emerald, turquoise and shimmering spring
Beckoning us to enter the majesty
To discover its secrets in the eerie depth."

TAKING ACTION

Getting rid of spartina won't be easy. It will take a lot of hard work and money to make a difference. That's why Vaughn and Guelker-Cone's 48 students feel the most important role they can play is educating parents, politicians and the public. Last month the class held town meetings and traveled to the state capitol in Olympia to talk about the spartina problem. When spring arrives, the kids will go to Padilla Bay to snip off spartina seed heads to keep the weed from spreading.

"One of the big lessons we learned from this project," says Vaughn, "was that we can't take care of the problem by just taking care of it in our bay; it is a problem for the entire Northwest coast."

The kids know that work to save the beauty of Padilla Bay has just begun. "We hope we can continue," says Guelker-Cone. It took decades for the spartina problem to take root; it will take many years to fix it.



Go On ►

1. Using the information from the article you just read, follow the instructions below to explain why spartina is a problem for the environment.

What is spartina?
Where does it come from?
Why was it was brought here?
How has the natural system been changed by the presence of spartina?

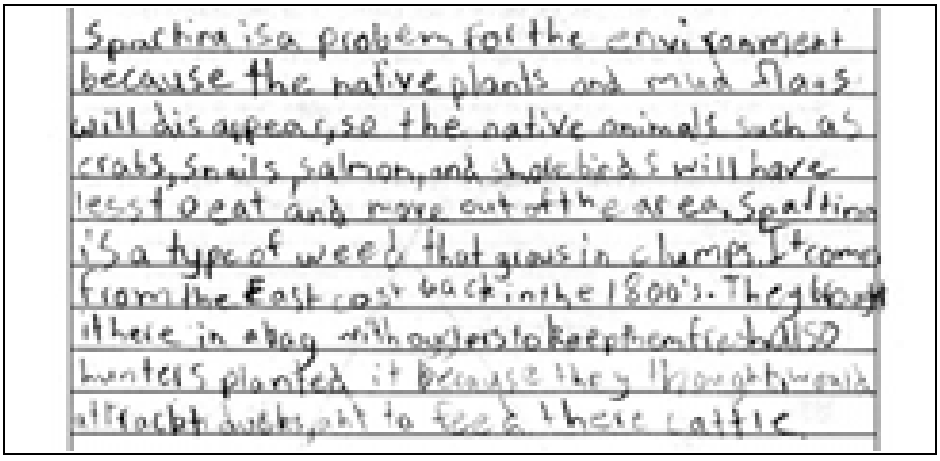
Go On ►

Scoring Rule Civics 8-G Question 1:

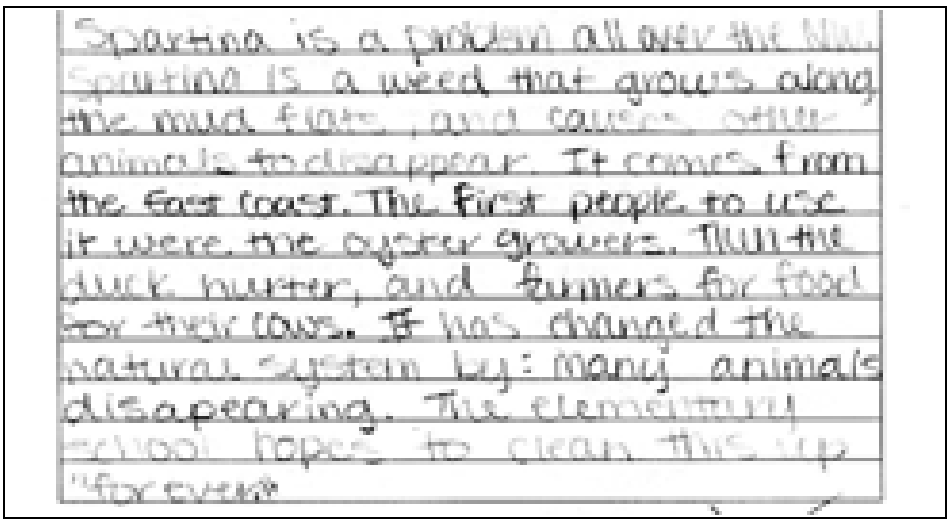
4	Response shows and excellent understanding of the how to extract information from text. Earns 7-8 value points: • Response describes what spartina is (1 value point), • Explains where it comes from (1 value point), • Provides three ways it was brought here (3 value points 1 point for each type of transport), • Provides three ways it has changed the natural system (3 value points, 1 point for each way it has changed the system).
3	Response earns 5-6value points.
2	Response earns 3-4 value points.
1	Response earns 2 value points.
0	Response shows no understanding of how to extract information from text; omitted; off-task.

Example Responses

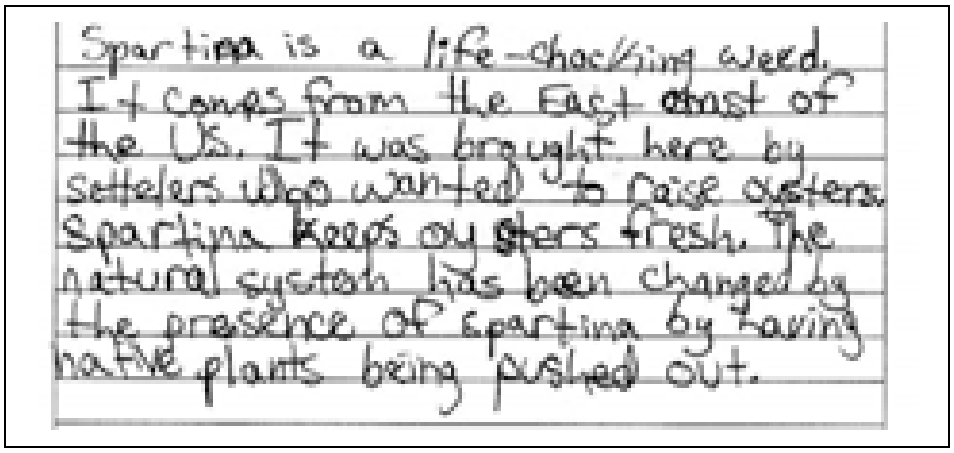
Student Work Civics 8-G Question 1 Score 4


Spartina is a problem for the environment because the native plants and mud flats will disappear, so the native animals such as crabs, snails, salmon, and shorebirds will have less to eat and more out of the sea. Spartina is a type of weed that grows in clumps. It came from the East coast back in the 1800's. They thought it here in bag with oysters to keep them fresh. Also hunters planted it because they thought it would attract ducks, but to see it here, cattle

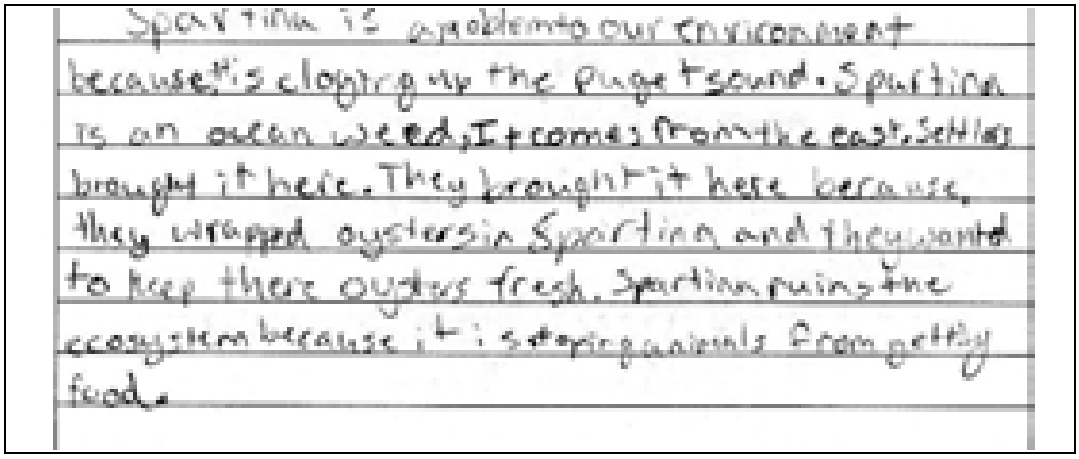
Student Work Civics 8-G Question 1 Score 3


Spartina is a problem all over the bay. Spartina is a weed that grows along the mud flats, and causes other animals to disappear. It comes from the East coast. The first people to use it were the oyster growers. Then the duck hunter, and farmers for food for their cows. It has changed the natural system by: many animals disapearing. The elementary school hopes to clean this up for ever.

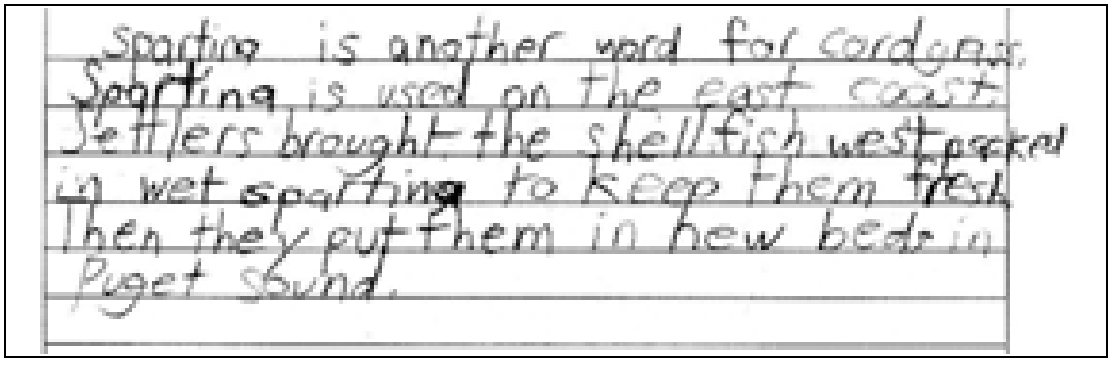
Student Work Civics 8-G Question 1 Score 2


Spartina is a life-chacking weed. It comes from the East coast of the US. It was brought here by settelers who wanted to raise oysters. Spartina keeps oysters fresh. The natural system has been changed by the presence of spartina by having native plants being pushed out.

Student Work Civics 8-G Question 1 Score 2


Spartina is a problem to our environment because it's clogging up the Puget sound. Spartina is an ocean weed. It comes from the east. Settlers brought it here. They brought it here because they wrapped oysters in Spartina and they wanted to keep there oysters fresh. Spartina ruins the ecosystem because it is storing animals from getting food.

Student Work Civics 8-G Question 1 Score 1


Spartina is another word for cordgrass. Spartina is used on the east coast. Settlers brought the shellfish westpacked in wet spartina to keep them fresh. Then they put them in new beds in Puget Sound.

2. What did the students do to make a difference?

Choose one action the students took and describe it
Explain how this action might help solve the spartina problem

Choose a second action the students took and describe it
Explain how this action might help solve the spartina problem

Go On ►

Scoring Rule Civics 8-G Question 2:

4	Response demonstrates an excellent understanding of how to identify and evaluate solutions to environmental issues. Earns 8 value points: • Describes one action (1 value points), • Describe second action (1 value points), • Fully explain how each action solves the problem (4 value points)
3	Response earns 6-7 value points
2	Response earns 4-5 value points
1	Response earns 2-3 value points
0	Response demonstrates little or no understanding of how to identify and evaluate solutions to environmental issues.; omitted off-task

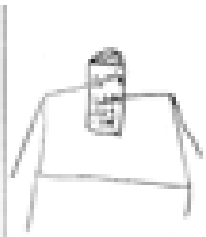
Example Responses

Student Work Civics 8-G Question 2:

Score 4

The students took many actions but here are two main actions. The 4th students and the 2 teachers set up town meetings and went to the state capital and they are going to go down to the bay and stop all spartina. Stop roads so they can't spread anymore. Setting up town meetings and going to the capital might help by getting people to help them get rid off spartina. And stopping of spartina and roads will stop them growing.

Score 4

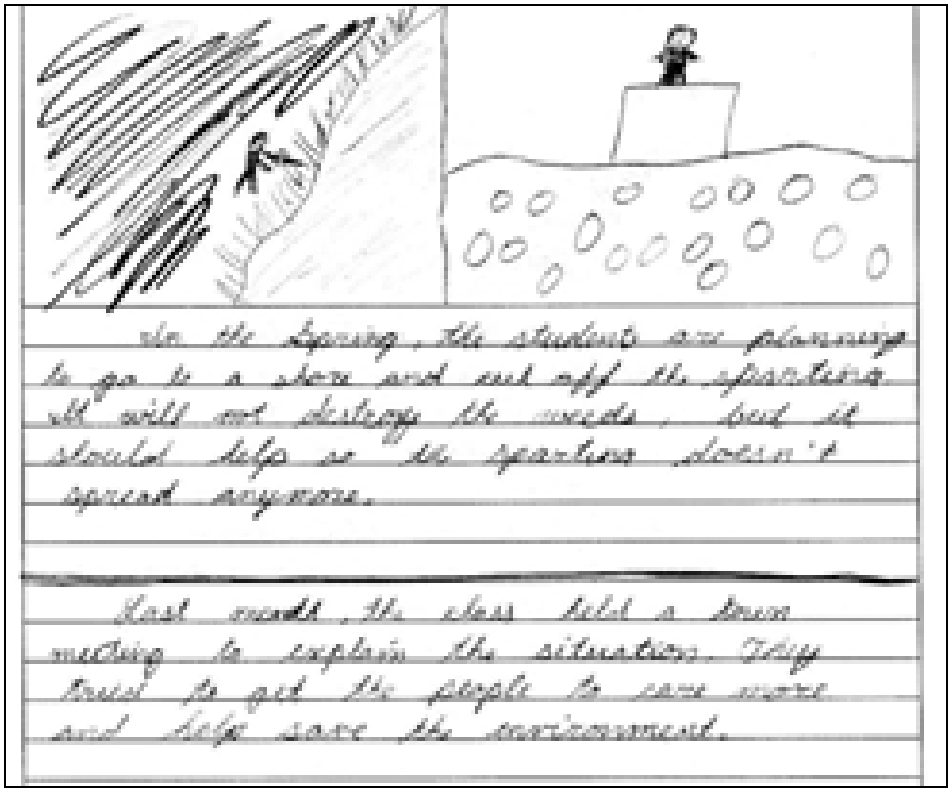


The students helped to make a difference by doing research on it. They looked in books and on the internet to find out as much info. on the plant as possible. They then contacted local experts. Then they found out where the plant grew (part sand, part rock) After doing all of the research they decided the during spring time they would go and cut off the Spartina and roots to prevent them from spreading. (the seed heads are when seeds are made, and also the wind blows they get spread to different places)

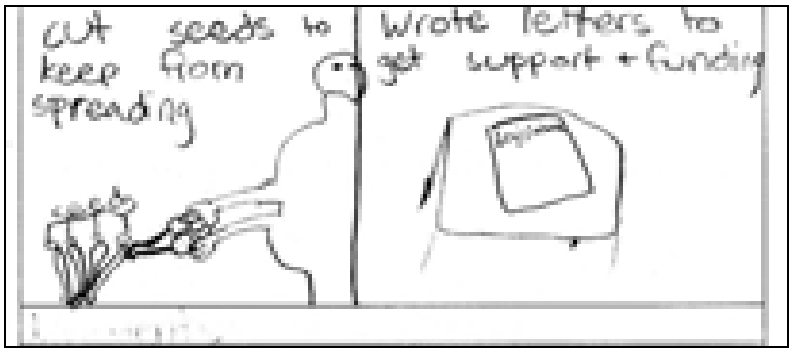
Student Work Civics 8-G Question 2 Score 3

The Students drew diagrams to show where the native plants were and to show where the spartina was, and the affect it had. This might help people see visually that the Spartina is bad.
Other students wrote letters to the stat legislature to try and tell them that the spartina is bad. This might help persuade the politicians to understand.

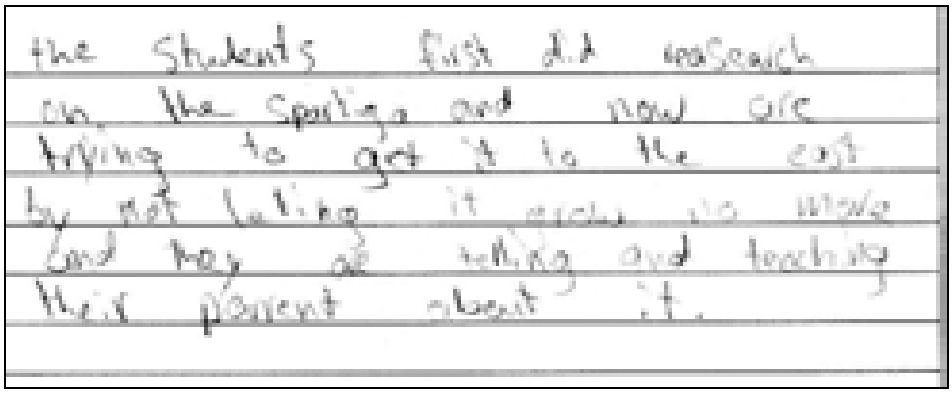
Student Work Civics 8-G Question 2 Score 3



Student Work Civics 8-G Question 2 Score 2



Student Work Civics 8-G Question 2 Score 1



3. Look at the poster the students made. Will the poster help others understand the problems created by spartina?

   - **Write a paragraph** explaining your answer to the question.
   - **Support your answer using information from the passage** and information from the poster.

Go On ►

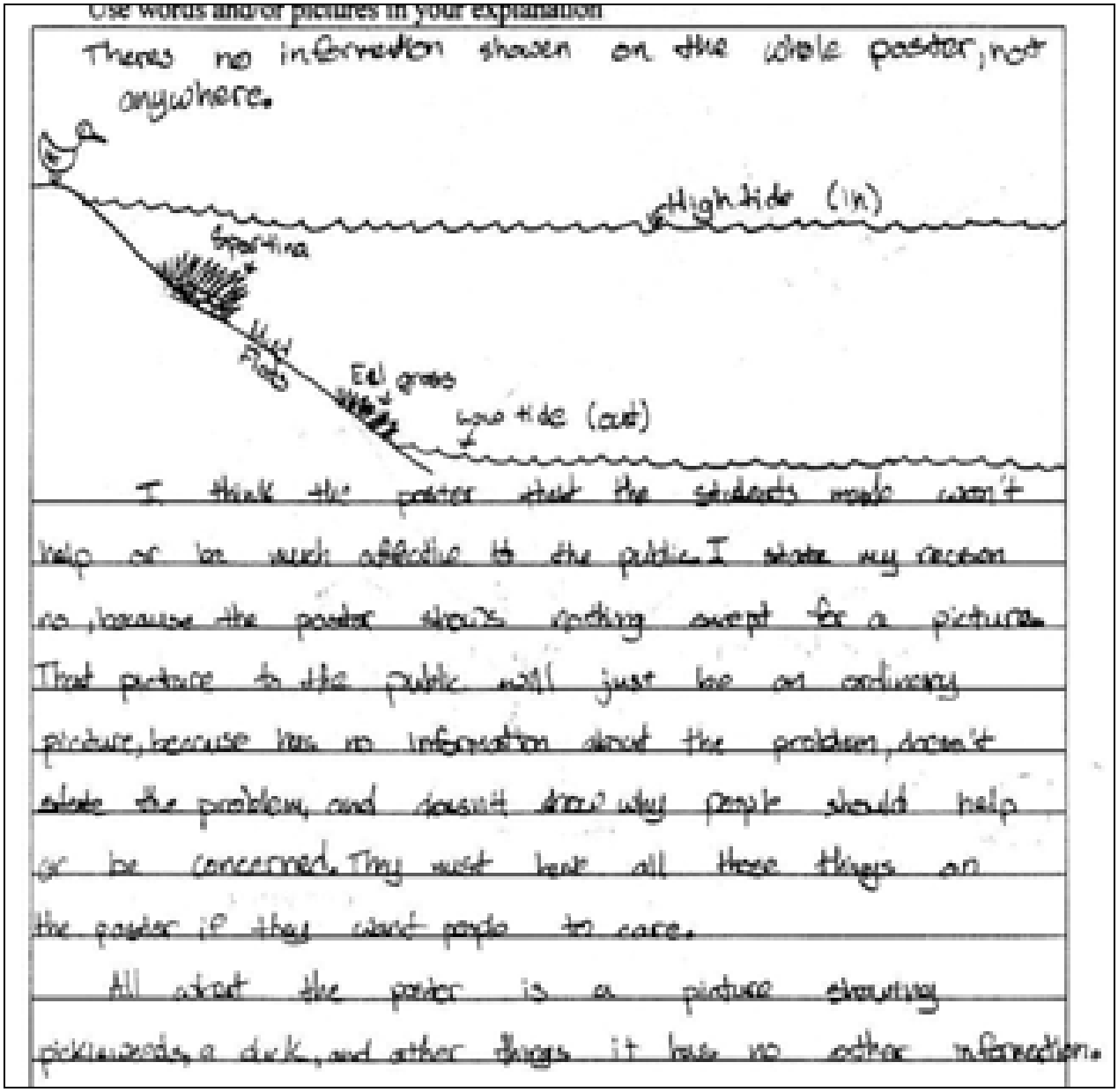
Scoring Rule Civics 8-G Question 3:

4	Response demonstrates excellent ability to critically evaluate the effectiveness of visual media and communicate findings. **Earns 8 value points**. **Writing Criteria** • Paragraph describes a complete thought (1 value point), • Paragraph is at least 2 sentences (1 value point) • Paragraph uses conventions of writing (1 value point), **Reading Criteria** • Accurately interprets and explains clearly how poster will or will not help people understand the problem (3 value points). Partially explains how poster will or will not help people understand the problem (2 value points), Determines whether people will or will not understand but the explanation is not clear (1 value point) • Supports answers with information from passage and poster (2 value points), supports answer with information from the passage or the poster (1 value point).
3	Response earns 6-7 value points.
2	Response earns 4-5 value points.
1	Response earns 2-3 value points.
0	Response demonstrates no ability to critically evaluate the effectiveness of visual media and communicate findings; omitted, off-task.

Example Responses

Student Work Civics 8-G Question 3:

Score 4

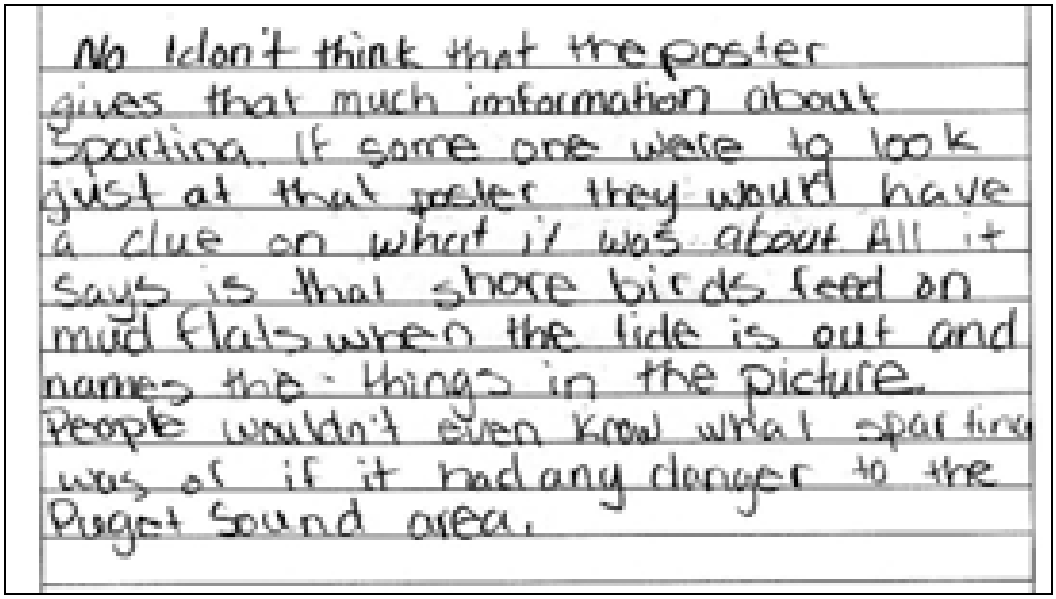

Use words and/or pictures in your explanation

There's no information shown on the whole poster, not anywhere.

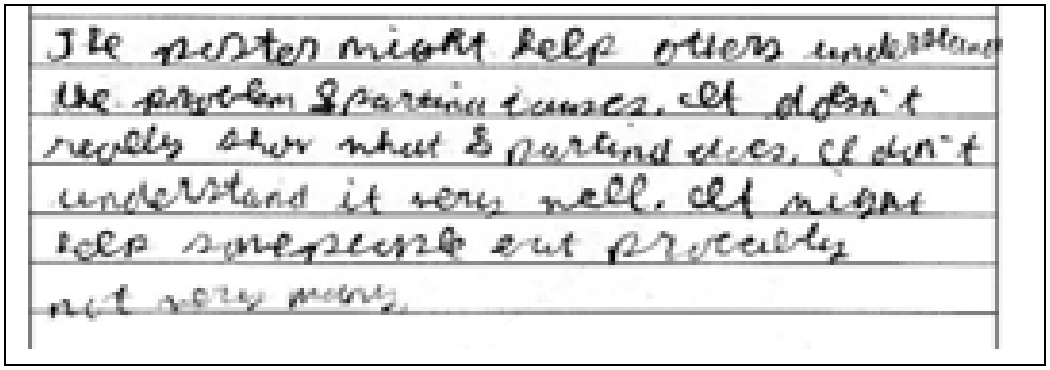
I think the poster that the students made won't help or be such effective to the public. I state my reason no, because the poster shows nothing except for a picture. That picture to the public will just be an ordinary picture, because has no information about the problem, doesn't state the problem, and doesn't show why people should help or be concerned. They must have all these things on the poster if they want people to care.

All about the poster is a picture showing pickleweeds, a duck, and other things it has no other information.

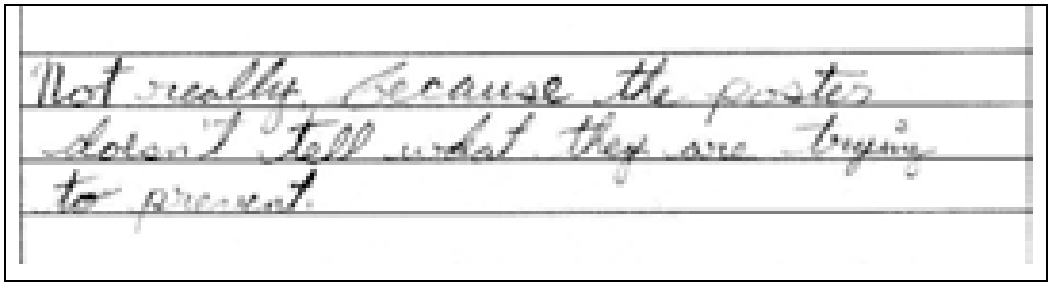
Student Work Civics 8-G Question 3 Score 3


No I don't think that the poster gives that much information about Spartina. If some one were to look just at that poster they would have a clue on what it was about All it says is that shore birds feed on mud flats wren the tide is out and names the things in the picture. People wouldn't even know what spartina was of if it had any danger to the Puget Sound area.

Student Work Civics 8-G Question 3 Score 2


The poster might help others understand the problem Spartina causes. It doesn't really show what Spartina does, it don't understand it very well. It might help someone out a little but not very much.

Student Work Civics 8-G Question 3 Score 1


Not really because the poster doesn't tell what they are trying to prevent.

4. In the space below, create your own poster about spartina. Use pictures and words on your poster.
   - Use information from the article to help you decide what message you want to give about the problems with spartina.
   - Decide how to share your message.
   - Make sure your poster will help people in Washington know what they can do to help take action.

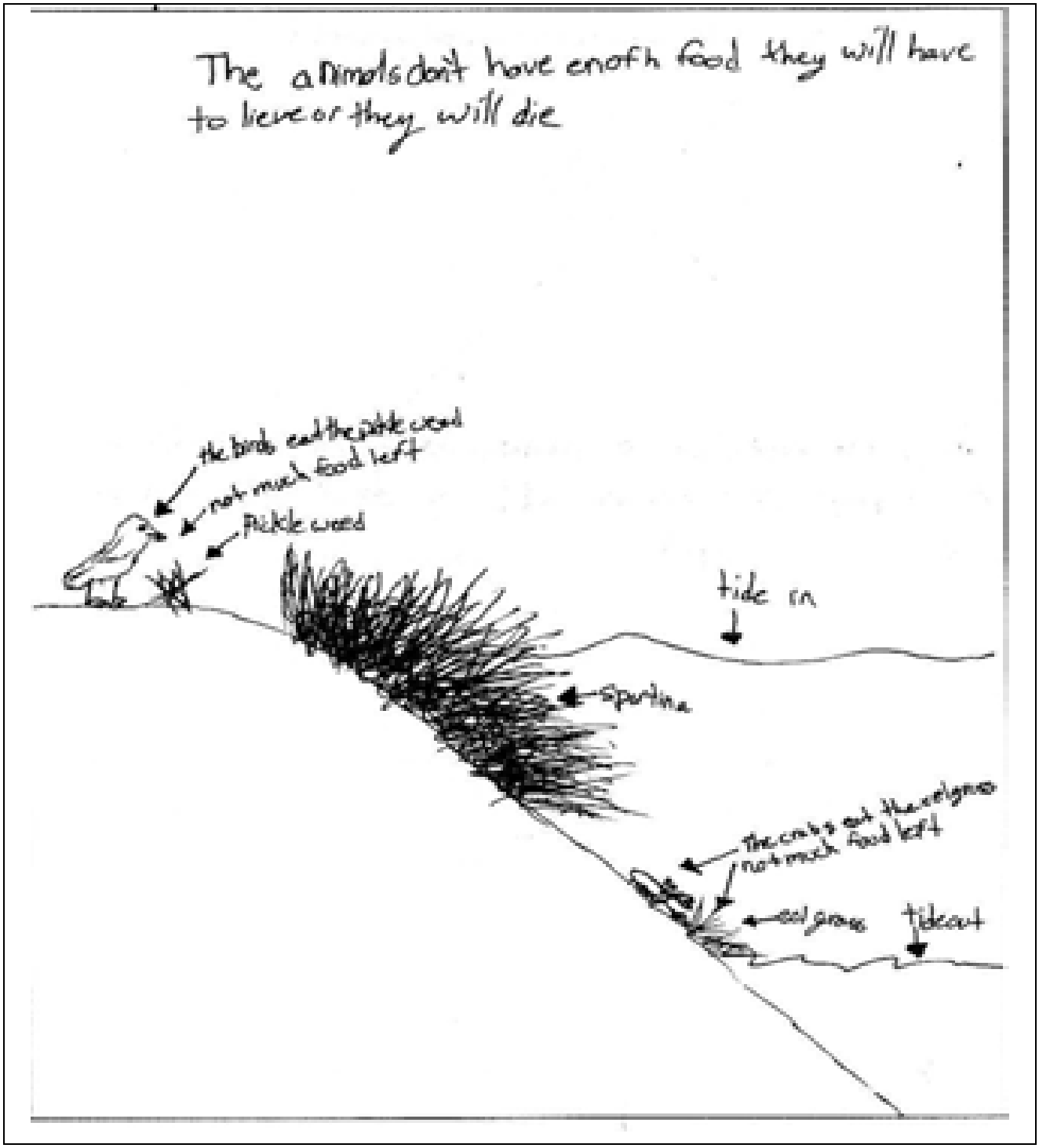
Go On ►

Scoring Rule Civics 8-G Question 4:

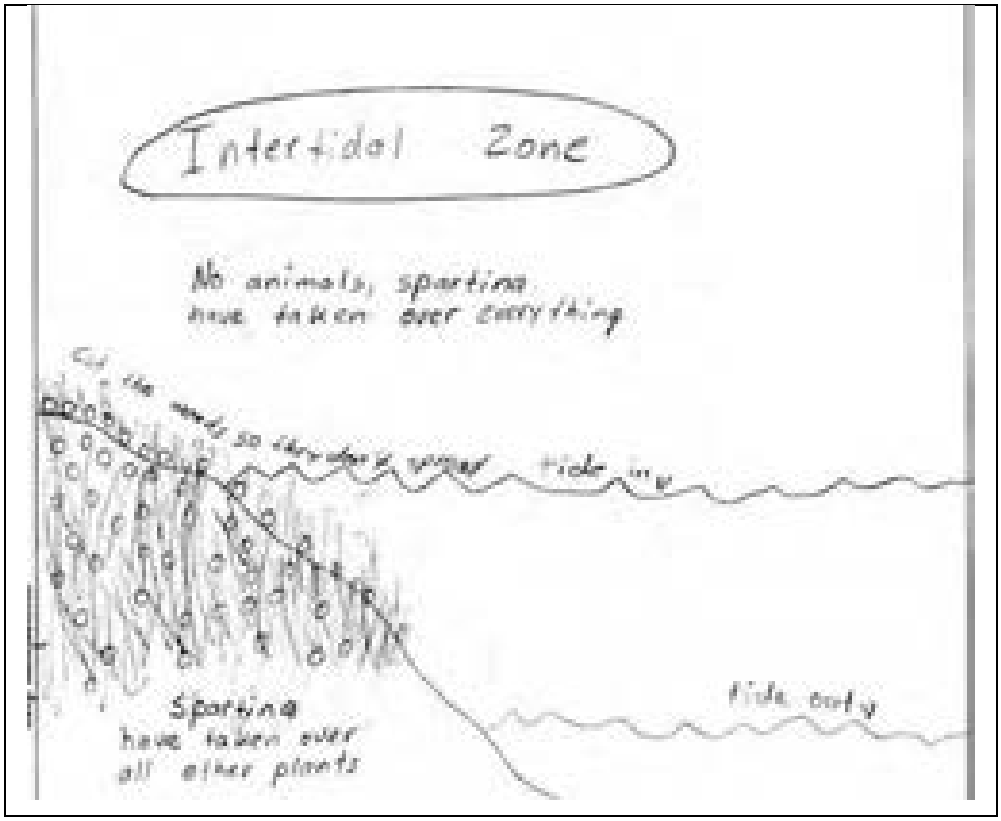
4	Response demonstrates excellent ability to express individual ideas and communicate effectively and responsibly. **Earns 6 value points**: Poster: • Demonstrates at least two elements of art (2 value points) demonstrates one element of art (1 value point) • Information describes problem (2 value points) Information partially describes the problem (1 value point) • Information describes a way the reader could take action to help fight spartina (2 value points)
3	Response earns 4-5 value points.
2	Response earns 3 value points.
1	Response earns 2 value points.
0	Response demonstrates no ability to express individual ideas and communicate effectively and responsibly; omitted; off-task.

Example Responses

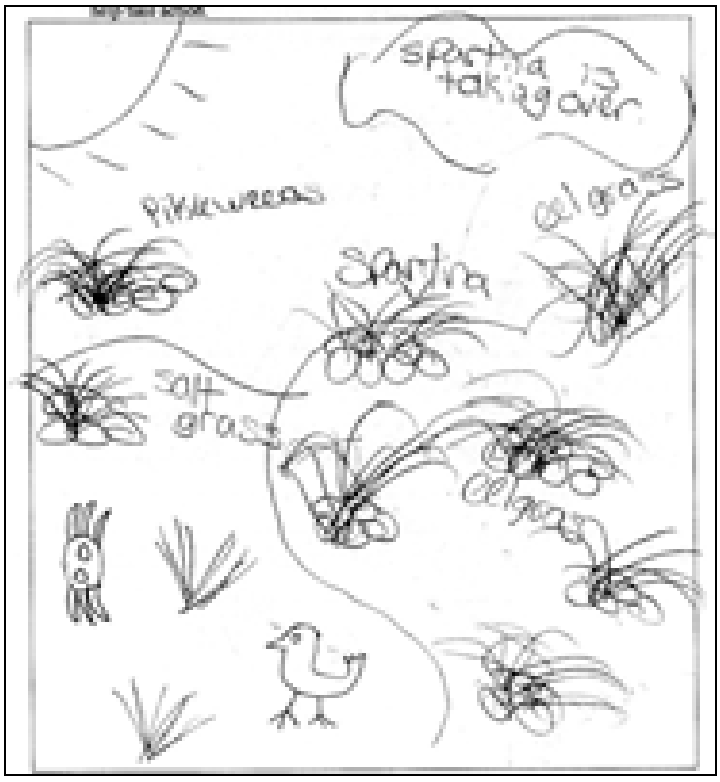
Student Work Civics 8-G Question 4 Score 4



Student Work Civics 8-G Question 4 Score 3



Student Work Civics 8-G Question 4 Score 2



An <u>action plan</u> is a plan for how to take action about an issue, a problem, or a concern. Think about your own neighborhood and community. What problem do you know about the environment in your neighborhood or town? In Numbers 5 through 9 you will begin an action plan.

5a. Describe one environmental problem in your neighborhood or town.
Explain how your neighborhood and the environment may be hurt by this problem.
Use words and/or pictures in your description and explanation.

Go On ►

5b. Write **two** questions that you will need to answer in order for you investigate the environmental problem in your neighborhood or town.

5c. Name **two** places you could go for information to answer your questions about the environmental problem in your neighborhood or town.

Go On ►

5d. Think of **one** action you could take to help solve the environmental problem in your neighborhood or town.

- Describe the steps in your action plan
- Explain how the action might help solve the problem.

Use words and/or labeled pictures in your description and explanation.

Go On ►

Before you can take action, it is important to consider how your actions might affect others.

5e. Identify **one** person or group that might be helped or hurt by your action plan.
Explain *why* the person or group might be helped or hurt.

5f. Identify **one** person or group you could talk to about your action plan to see if it will work.

Explain why the person or group could help you with your plan.

STOP

Go back through the booklet to make certain you are finished with your work. Then give your booklet to your teacher.

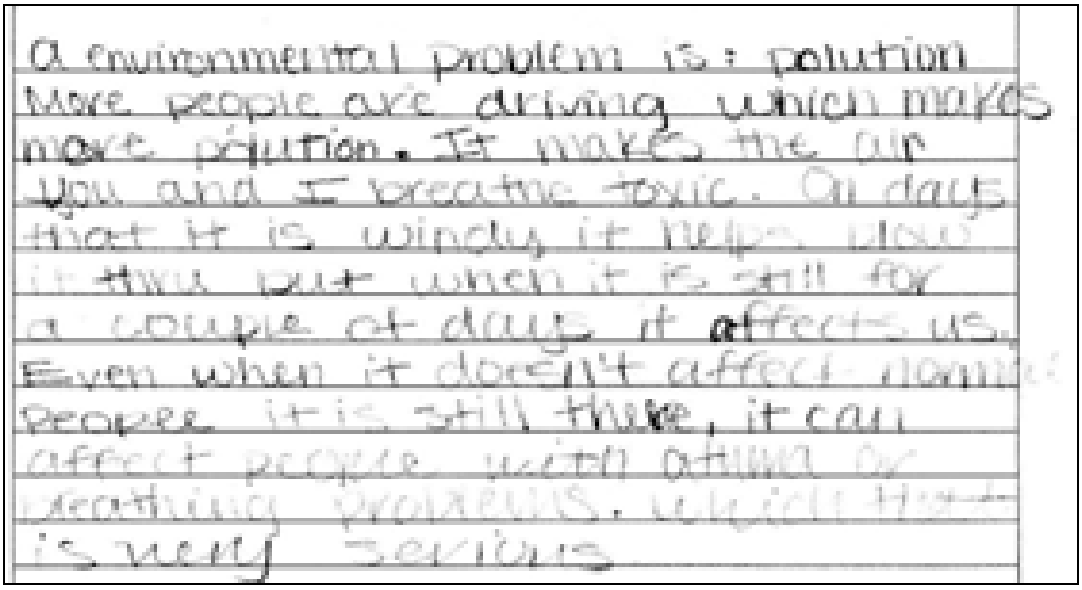
Scoring Rule Civics 8-G Questions 5:

4	Response shows excellent understanding of the steps necessary to implement an action plan. **Earns 14-16 value points**: Action Plan response: 5a - identifies an environmental problem (1 value point) and explains how the problem could hurt their neighborhood/environment (2 value points) 5b - selects two questions that need to be investigated that relate to the problem (1 value point for each question) 5c - identifies two plausible places to get information (1 value point for each place), 5d -selects one action (1 value point) with at least three steps necessary to implement the plan (2 value points) identifies 1- 2 steps (1 value point) and explains how the steps will help (2 value points) 5e-Identifies **one** person who would be helped/hurt by the plan (1 value points), and **explains why** the person would be helped or hurt (1 value point) 5f-Identifies the name of **one** person to talk to see if it will work (1 value points), and **explains why** the person or group could help with the plan (1 value point)
3	Response earns 10-13 value points
2	Response earns 6-9 value points
1	Response earns 3-5 value points
0	Response shows no understanding of the relationship between human activity and natural systems; omitted; off-task.

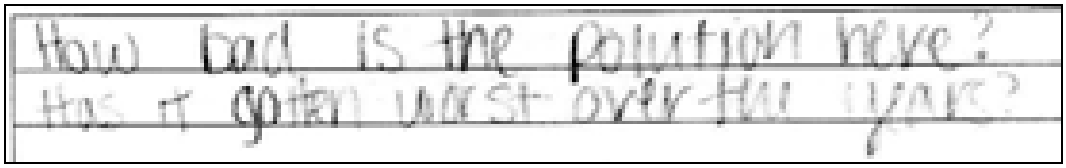
Example Responses

Student Work Civics 8-G Questions 5 Score 4

5a-contributes 3 value points


A environmental problem is: polution
More people are driving which makes
more polution. It makes the air
you and it breathe toxic. On days
that it is windy it helps. How
it flows but when it is still for
a couple of days it affects us.
Even when it doesn't affect many
people it is still there, it can
affect people with asthma or
breathing problems. Which this
is very serious

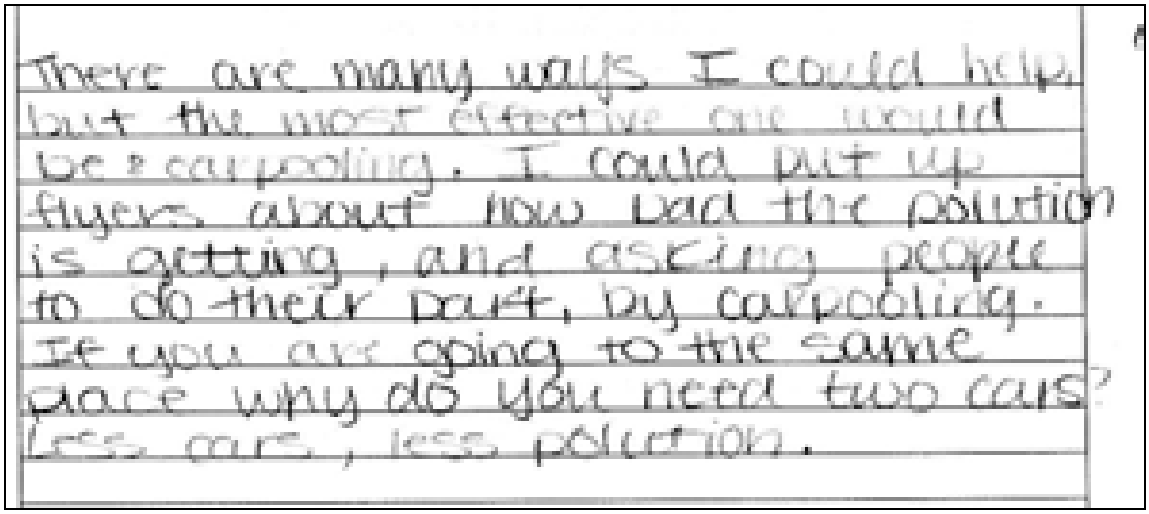
5b -contributes 2 value points


How bad is the polution here?
has it gotten worst over the years?

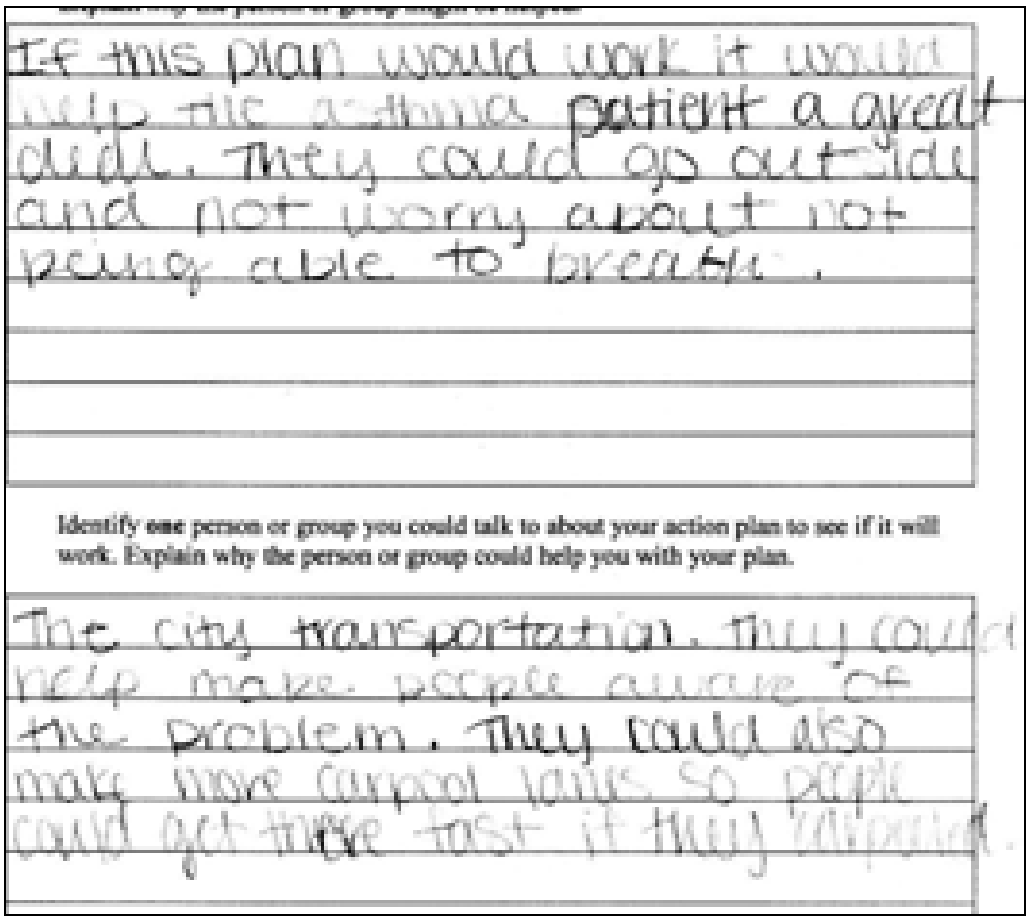
5c-contributes 2 value points


Library & Meteorologist

5d – contributes 4 value points


There are many ways I could help
but the most effective one would
be carpooling. I could put up
flyers about how bad the polution
is getting, and asking people
to do their part, by carpooling.
If you are going to the same
place why do you need two cars?
less cars, less polution.

5e and f- Contributes 2 value points



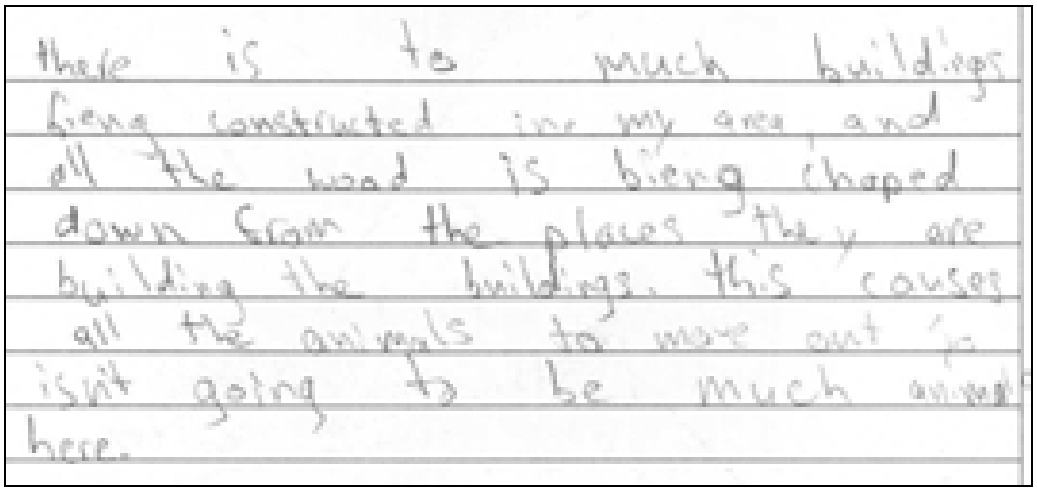
If this plan would work it would help the asthma patient a great deal. They could go outside and not worry about not being able to breath.

Identify **one** person or group you could talk to about your action plan to see if it will work. Explain why the person or group could help you with your plan.

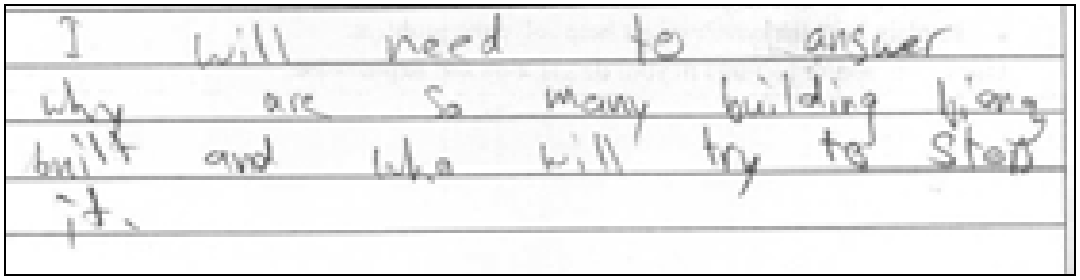
The city transportation. They could help make people aware of the problem. They could also make more carpool lanes so ppl could get there fast if they carpool.

Student Work Civics 8-G Questions 5 Score 3

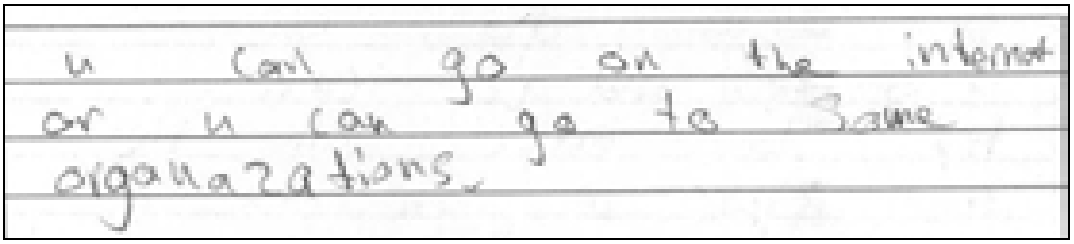
5a-contributes 3 value points



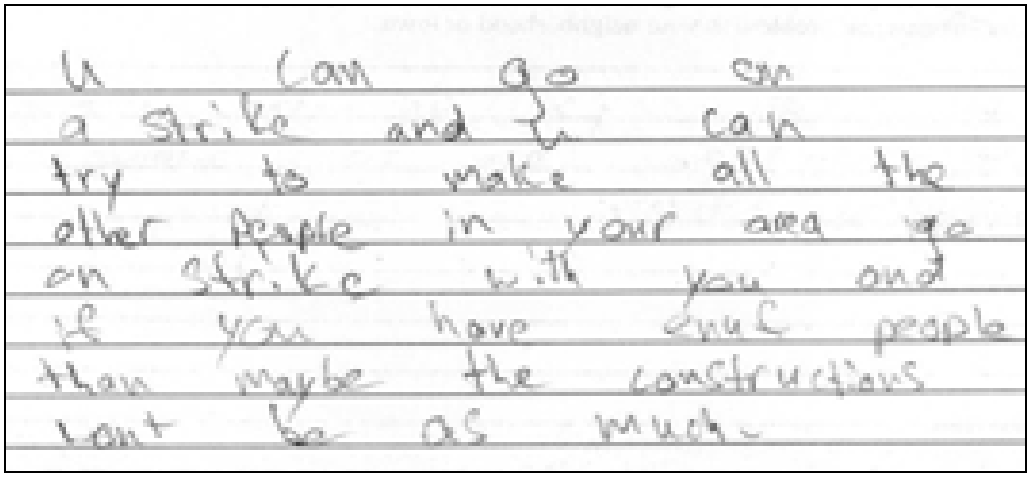
5b-contributes 2 value points



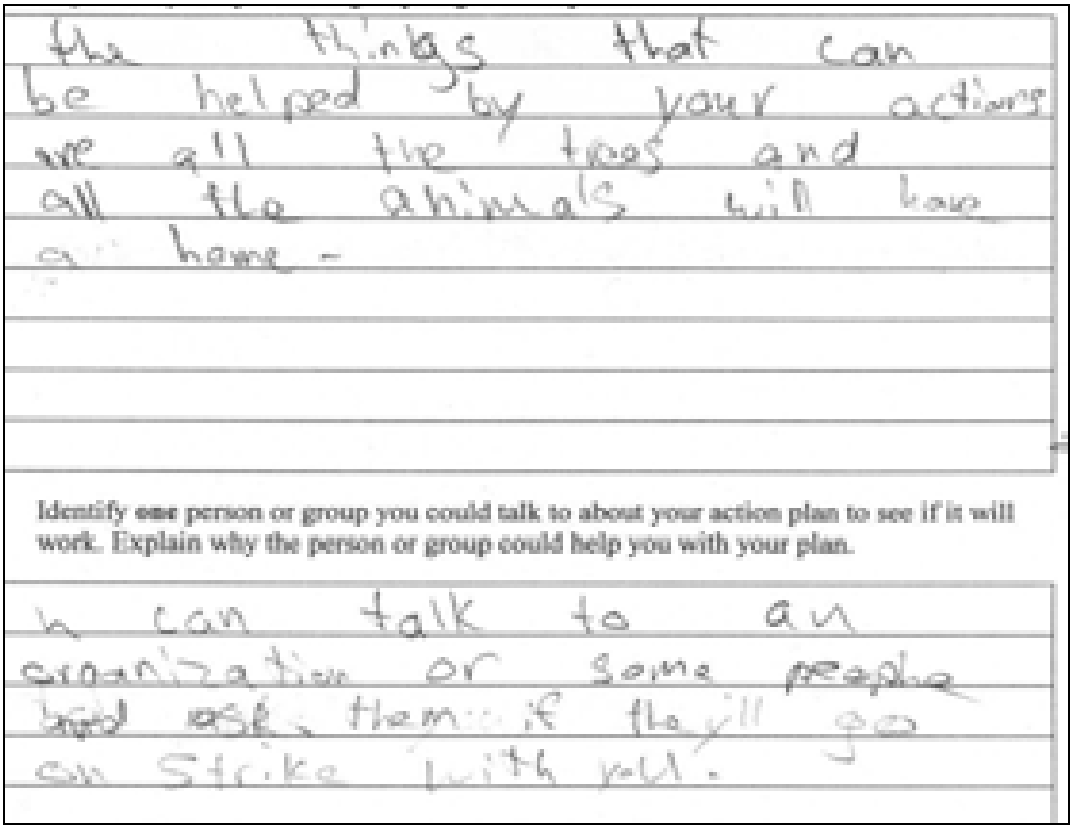
5c-contributes 1 value points



5d-contributes 3 value points



5e&f contributes 2 value points



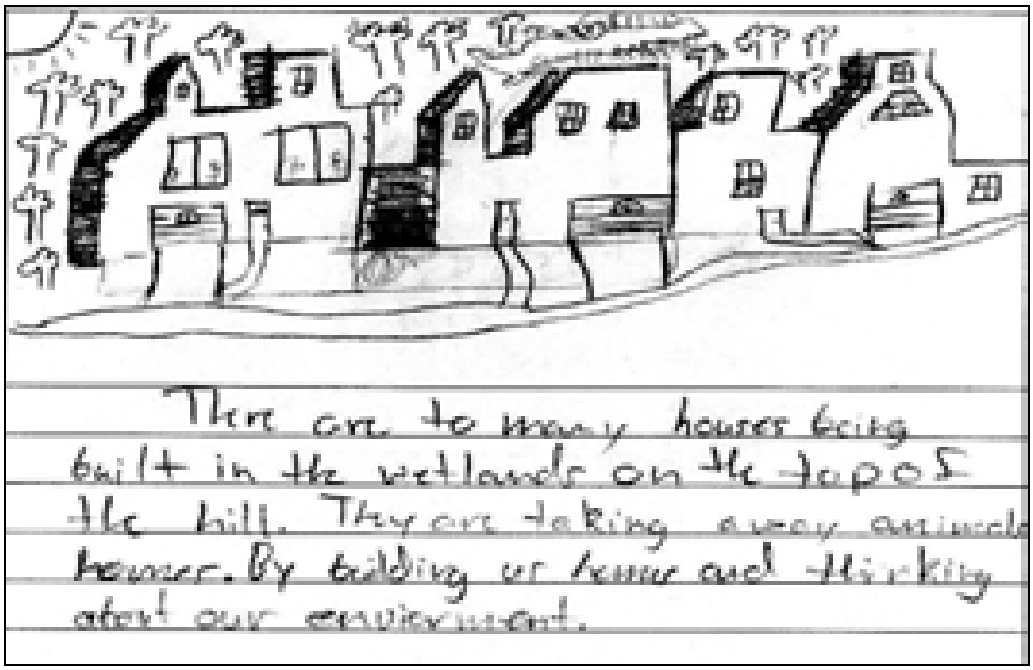
the things that can be helped by your actions we all the trees and all the animals will have a home -

Identify one person or group you could talk to about your action plan to see if it will work. Explain why the person or group could help you with your plan.

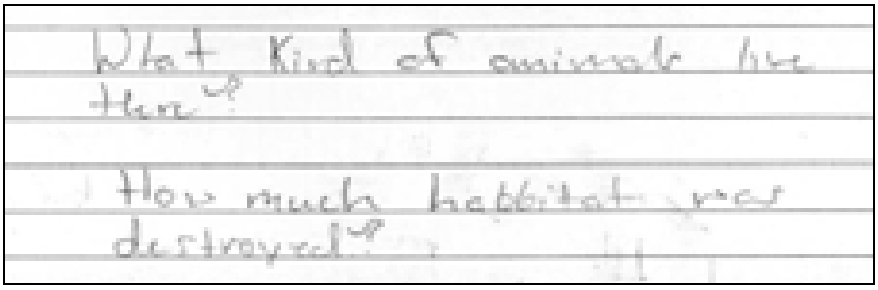
I can talk to an organization or some people and ask them if they will go on strike with you.

Student Work Civics 8-G Questions 5 Score 2

5a-contributes 2 value points



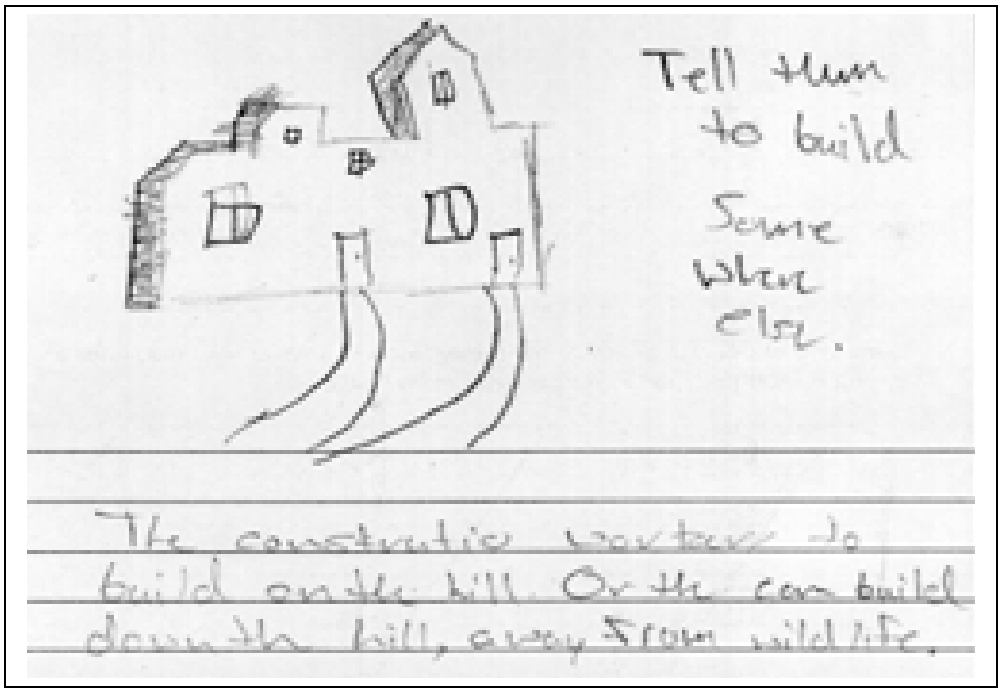
5b- contributes 2 value points



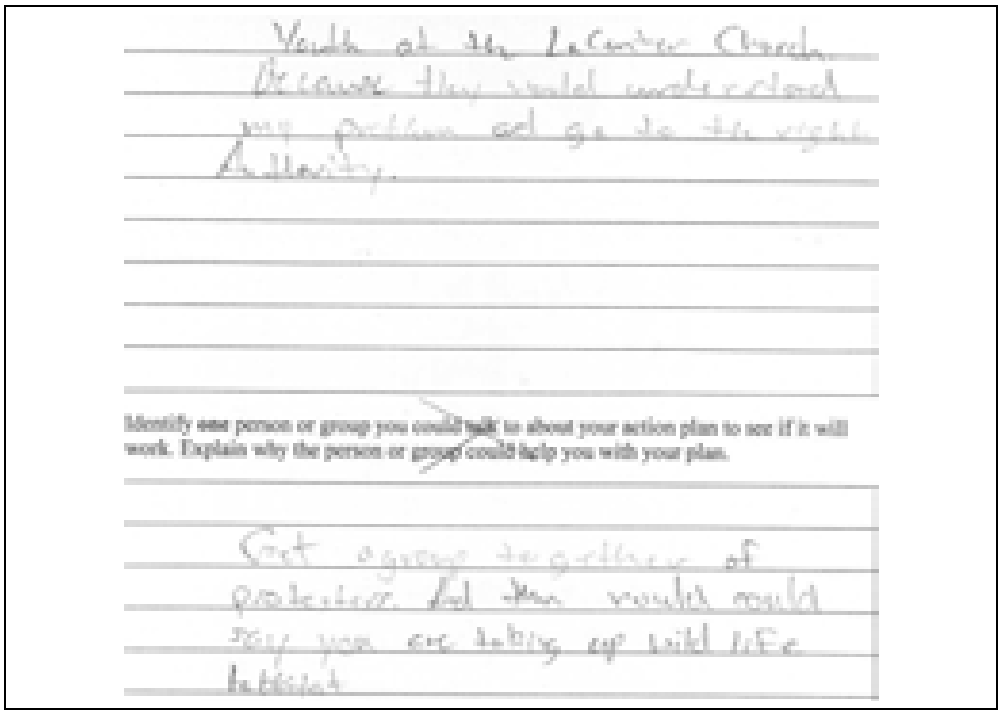
5c-contributes 0 value points

No response

5d-contributes 1 value point



5e&f-contributes 0 value points



Appendix C

Project Development Partners for the Environmental Education Assessment Project (PEI)

Primary Environmental Education Programs:

Washington Forest Protection Association -Project Learning Tree (Lynne Ferguson, Kathryn Kurtz Smith, Pat Otto, Woody Franzen, Susan Duncan); **Washington Department of Fish & Wildlife -Project Wild** (Margaret Tudor, Nicole Rickert); **Dept. of Educational Psychology University of Washington,** (Dr. Catherine Taylor), **Office of the Superintendent of Public Instruction** (Tony Angell, Brenda Hood) ; **Center for Ecogenetics and Environmental Health, University of Washington**(Katie Frevert, Jon Sharpe); **Washington School Principals' Educational Foundation and Cispus Environmental Learning Center** (Marty Fortin); **National Project Learning Tree** (Kathy McGlauflin); **State Environmental Education Roundtable** (Jerry Lieberman, Linda Hoody), **Washington State Department of Natural Resources** (Barbra MacGregor), **Washington State Audubon** (Heath Packard, Naki Stevens, Beth Doglio).

Supportive Environmental Education Programs: **Center for Environmental Education, WSU** (Kim Frier); **Washington Department of Ecology -Project WET** (Rhonda Hunter,) ; **North Cascades Institute** (Tracey Johanneson); **Olympic Park Institute** (Lisa Eshenbach); **Weyerhaeuser's Environmental Education Initiative** (Jim Stark, Martha Avery, Katie Taylor); **National Park Service** (Lisa Eschenbach, Mike Dedman); **Tahoma Audubon** (Katrina Weihs); **Woodland Park Zoo** (Dave Hill, Nicole Rickert);
Colleges and Universities: **Central Washington University** (Martha Kurtz), **Eastern Washington University** (Roger Hauge), **Gonzaga University** (Jonas Cox, Anne Martin), **St. Martin's College** (Paul Nelson), **Whitworth College** (Dennis Sterner)**, University of Puget Sound, Seattle University, The Evergreen State College, University of Washington** (Catherine S. Taylor, Mark Windshitl), **Washington State University** (David Gruenwald, Virginia Newhouse-Rogers, Jeff Sellen)**, Seattle Pacific University** (Ray Myers)**, Western Washington University, Heritage College** (Apanakhi Buckley).

School Districts: **Issaquah** (Phyllis Runyon, Lynne Vona, Leslie Smith, Jane Ulrich, Barbara Walton, Nancy Drake); **Tahoma** (Nancy Skerritt, Kristin Edlund, Ethan Smith, Bud Cross, Jerry Papers, Robert Richter); **Tumwater** (Sue Haskin, Laurie Westman, Art Hoover, George Rother, Erica Baker, Marchand Connelly, Pat Lisoske), **Peninsula** (Sally Gallagher, Sylvia Wallen, Dennis Nugent), **Kent** (Jeannie Mong, Jon Wilcox), **Bainbridge Island** (Marie Marrs, Jim Corsetti); **Battle Ground** (Diane Townsend), **Cle Elum** (Sharon Maras, Sue Gribble, Katie Gray, Ryan Hill, Trish Griswold, Connie Anderson); **West Valley** (Tom Moore, Katie Owens, Sue Fisher, Jamie Ostby, Eric Groshoff), **Everett** (Kayleen Pritchard, Nan Lombardo, Liz Wilson, Suzanne Cowper, Gary Whitney) , **Port Angeles** (Sandra Biasell), **Shelton**, **Tacoma** (Marlene Rossi, Laurel Mosher, and entire staff Whitman Elementary), **Puyallup** (Karen Saxon); **East Wenatchee** (Dave Speis); **Lake Washington** (Eileen McMakin, Brian Healy, John), **Elma** (Jeff Dunn, Kathleen Pettel-Price); **Edmonds** (Sonja Bickell); **Deer Park** (Carolyn Henry); **Evergreen** (John Akers)

Educational Service Districts: **#112**, **Vancouver; #105, Yakima** (Marsha Pastrana)**; #123, Pasco** (Diane Greenwald, Carol Pacheo, Mary Rosier)**; #189, Mount Vernon** (Lonnie Pithan)**; #113, Olympia** (Sue Shannon)**; # 101, Spokane** (Helene Paroff)**; #114** (Dennis MCrea)

Schools: **Environmental Adventure School** (Lake Washington School District), **McLane Elementary** (Olympia School District).