

# Performance-Based Task Assessment Of Higher-Order Proficiencies In Redesigned STEM High Schools

Jeremy V. Ernst, Virginia Tech, USA  
Elizabeth Glennie, RTI International, USA  
Songze Li, Virginia Tech, USA

## ABSTRACT

*This study explored student abilities in applying conceptual knowledge when presented with structured performance tasks. Specifically, the study gauged proficiency in higher-order applications of students enrolled in earth and environmental science or biology. The student sample was drawn from a Redesigned STEM high school model where a tested performance assessment protocol was employed for the purposes of the investigation. It was determined that performance-based proficiency was not uniform within tasks and applications, but could be recognized through student artifacts of learning on a situational basis. Based on the findings of the study, several implications are highlighted.*

**Keywords:** Performance-based Assessment; Higher-order Proficiency; Redesigned STEM Schools

## INTRODUCTION

Performance-based assessments require students to engage in certain activities or create products to demonstrate their academic knowledge and abilities. Tasks in performance-based assessments are closely related with procedural knowledge (Alsardary, Pontiggia, Hamid, & Blumberg, 2011). Procedural knowledge, different from declarative knowledge, is usually gained through observation of people's actions. Assessing procedural knowledge often involves problem-solving tasks (Anderson et al., 2001). When intentionally targeting procedural knowledge measurement, performance-based assessments typically seek student responses in alternative types (such as short answer, essay, drawing, and data tabulation) other than multiple-choice selection, in order to document and reflect what students do and how they reason and formulate conclusions.

Six cognitive processes have been defined in Revised Bloom's Taxonomy (Anderson et al., 2001): remember, understand, apply, analyze, evaluate, and create. The latter four processes are usually considered higher-order thinking. In contrast to the conventional standardized tests, which commonly focus on content knowledge, performance tasks usually employ a variety of activities targeting higher-order proficiencies (Pinter, Matchock, Charles, & Balch, 2014). Research shows that performance-based assessments could help monitor a set of complex learning objectives, such as reasoning and problem-solving abilities, in various subject contexts (Hambleton & Murphy, 1992; Rudner & Boston, 1994). Incremental validity and reliability of performance-based tests over standardized tests have also been found in various educational contexts across grade levels (Morton, Cumming, & Cameron, 2007; Tanihon, Segers, Vedder, & Tillema, 2009; Falk, Wichterle Ort, & Moirs, 2007).

Performance-based assessments are conducive to contemporary educational theories, frameworks and standards. Constructivists believe that learning occurs when people construct meaning of the engaged activities. While assessing student competence, performance tasks offer students with authentic experiences through which new knowledge can be constructed. Hence, performance-based assessments administered to students can function as an instructional tool, and potentially enhance students' learning (Falk, Wichterle Ort, & Moirs, 2007).

The recently released Next Generation Science Standards (NGSS) emphasize eight “practices” in K-12 science standards. Students’ performance is considered as an indispensable way to demonstrate competency on knowledge and skills specific to each practice in science learning:

*Standards and performance expectations that are aligned to the framework must take into account that students cannot fully understand scientific and engineering ideas without engaging in the practices of inquiry and the discourses by which such ideas are developed and refined. At the same time, they cannot learn or show competence in practices except in the context of specific content.* (National Research Council, 2012, p. 218)

The NGSS requires science assessments to incorporate both students’ understanding of core ideas and their abilities to use the practices of science and engineering. Under this new expectation, students should be assessed on not only the understanding and applying of factual knowledge, but also higher-order abilities of scientific investigation and engineering/technology design. Performance tasks offer such integration of knowledge, skills, and abilities (Nitko & Brookhart, 2015).

Students, especially underserved and underrepresented students in science (such as unprivileged minorities, students enrolling in free or reduced-price lunch), are receptive to such experiential educational opportunities (Oberg, 2009). Student-centered approaches are grounded in the cognitive and pedagogical research of the past decades (McLoughlin & Taji, 2005). This research advocates to realize the importance of the cognitive dimension in learning and to support student cognitive development through instruction and verification through appropriate assessment. Although performance practices and assessments will not negate the differences among cultural or socioeconomic groups, student-centered instrumentation design and administration have the potential to promote the equitable use of performance-based assessment (Darling-Hammond, 1994). Performance assessments, when coming from a student-centered perspective, allow multiple representations or formats in expression of student’s knowledge and abilities. Therefore, students will have a level of flexibility in response selection based on their unique backgrounds and prior experiences.

### **PERFORMANCE-BASED APPLICATIONS**

Performance-based learning, as defined by Voorhees (2001), is “learning systems that seek to document that a learner has attained a given competency or set of competencies (as cited in Cydis, 2015, p. 70). As a result of competency-oriented learning practice, performance-based learning activities have to coordinate with target learning objectives. The performance tasks and assessment techniques have to be complimentary and in alignment to these stated objectives.

Unlike conventional summative standardized tests, performance-based assessment is integrated with the whole learning process as learning is embedded in the actual assessment tasks. Performance tasks serve as both an integral part of learning activity and an opportunity to assess the learning outcomes (Hibbard, 1996). Students develop a meaningful connection to the content and construct new knowledge when engaged in the performance tasks (Cydis, 2015).

The scoring of performance-based assessment should reflect the capabilities of students rather than the rater’s perceptions and biases (Stiggins, 1987). Therefore, a consistent, reliable scoring system is critical to the fairness of performance-based assessment. Among various scoring techniques, scoring rubrics, which describe the characteristics of different levels of performance, have been accepted as a predominant tool of performance assessment (Kan, 2007). Scoring rubrics can include both quantitative and qualitative description on performance criteria (Mertler, 2001). Therefore, it is more effective and suitable than conventional standard-answer scoring for evaluating student cognitive abilities in higher-order cognitive dimensions.

Two types of scoring rubrics, holistic and analytic, have been commonly used in assessment. The former provides an overall score of the process or product directly, while the latter scores individual components separately to obtain a collective score (Nitko & Brookhart, 2015). When scoring student STEM performance based on competency-based learning objectives, an analytic rubric would be more appropriate to address each attribute.

Generally, performance-based assessment is measured by means of observation and professional judgment (Stiggins, 1987). Compared to many other assessment techniques, performance-based assessment typically require additional effort from teachers in design, administration, and especially grading. However, Arter (1998) suggested that the complexity should match between the learning objectives and the assessment when choosing methods of measuring learning outcomes. In order to address the complex learning objectives on science concepts and practices required by state standards, performance-based assessment is needed to analyze students' knowledge application and higher-order thinking skills.

### RE-DESIGNED STEM SCHOOLS

Supported by the Bill and Melinda Gates Foundation and the North Carolina General Assembly and the North Carolina Department of Public Instruction, educators from local and higher education have made a collaborative effort to create innovative STEM high schools. North Carolina New Schools (NCNS) started in 2007 establishing STEM schools to function as laboratories for students to solve real-world problems using technology, understand relevance among science, technology, and mathematics, and experience out-of-school learning in co-curricular activities (NCNS, 2013). These schools are intentionally designed to have small class sizes, with 100 students in each freshman class. The common instructional framework designed for these schools emphasizes collaborative group work, writing to learn, questioning, scaffolding, classroom talks, and literacy groups in all classes, in order to enhance student exploration and invention and to foster a culture of collaborative inquiry (NCNS, 2014).

Five schools, varying in locals, ethnicity, and poverty compositions, participated in the performance assessments in this study. They came from two categories: Some were schools that were redesigned as Science or Technology schools; the others were early college high schools located on college campuses, offering college classes and associate's degree or two year transferable credits to the University of North Carolina System. The school characteristics about STEM type and student compositions are summarized in Table 1. Student performance in the school year of 2012-13 in Algebra 1 and Biology as well as the teacher's qualifications are shown in Table 2 to provide background information about school academic performance.

**Table 1.** School Characteristics

Site	Urbanicity	STEM Type	Total Students	Percent Under-Represented Minority	Percent Free or Reduced-Price Lunch
1	Rural	Early college	160	61.3	65.0
2	City	Science high school	188	71.3	47.9
3	City	Early college	257	49.8	32.3
4	Town	New Tech high school	148	57.4	70.3
5	Rural	Early college	275	19.6	44.4

**Table 2.** Student Performance and Teacher Qualifications in STEM Schools

Site	Percent passing Algebra 1 End of Course exam	Percent passing Biology End of Course exam	Percent fully-licensed teachers	Percent novice teachers
1	63.4	74.3	100	50.0
2	35.3	64.6	93.8	12.5
3	64.3	84.5	100	13.3
4	0.0	22.2	100	25.0
5	63.6	77.3	100	14.3

## RESEARCH QUESTIONS

In this study, researchers investigated student's performance proficiency in science activities for these re-designed STEM high schools. The investigation was guided by the following research questions:

- 1) To what extent do students in re-designed STEM high schools demonstrate identifiable proficiency in performance-based earth/environmental science assessments?
- 2) To what extent do students in re-designed STEM high schools demonstrate identifiable proficiency in performance-based biology science assessments?

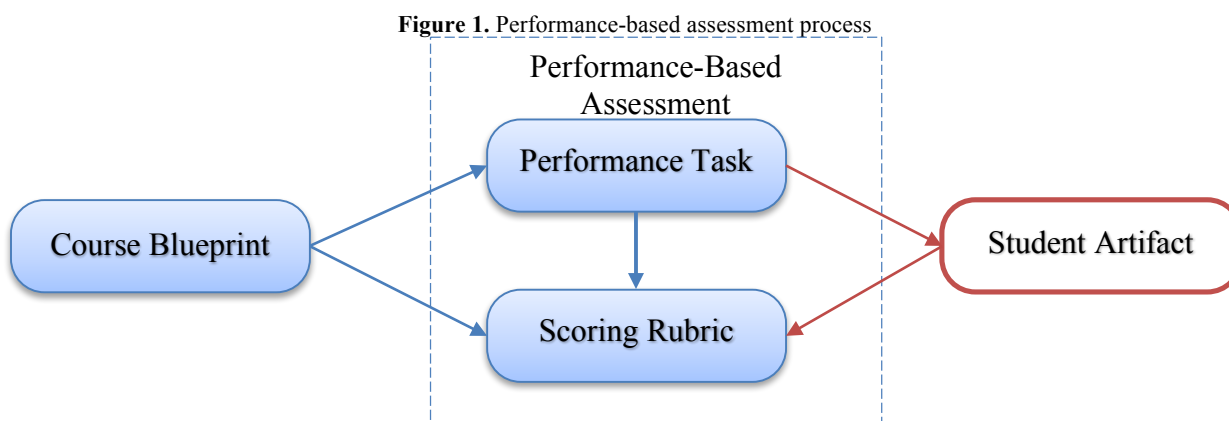
These proficiencies, including evaluation, prediction, analysis, synthesis, and reasoning, were examined in different contexts for both earth/environmental science and biology. For earth/environmental science, the contexts include soil and water, clouds and weather, acid rain, earthquake and seismometry, and astronomy-based performance tasks. For Biology, the performance tasks are assigned in the contexts of inheritance and molecular genetics, cell organelles, photosynthesis and cellular respiration, human activities and the environment, and enzyme and fruit browning.

## METHODOLOGY

The research methods in the study closely followed Ernst & Glennie's (2015) research. During the school year when the performance assessments were administered, high school students in North Carolina were required to take three science courses including Earth/Environmental Science and Biology (North Carolina Department of Public Instruction [NCDPI], 2012a). The New Tech STEM High Schools and Early College High Schools were contacted for participation request and five schools responded positively. Three Earth/Environmental Science and two Biology classes, based on the regular course offering in these five schools, were involved in the study.

Project parental consent and student assent forms were distributed to the participating students and their parents prior to the onset of the study. Sixty-three signed forms with student products were collected in this study during school year 2013-14. A set of performance assessments in either Earth/Environmental Science or Biology (see Instrumentation section) was provided to the five participating teachers. Teachers received step-by-step instruction, scoring rubrics, and blank student notebooks. Materials and instructional support were offered to teachers upon their request. Materials and tools in each activity of the assessments was limited to a standard teacher inventory. The teachers were expected to have their students conduct five performance tasks as supplemental activities to their regular instruction.

These performance tasks and corresponding scoring rubrics directly addressed competencies specified in state standards and required application of course content and exploration. Students were expected to document their performance process, research findings, reflections, and proposed solutions in the provided notebook as guided by the assessment tasks. These notebooks were collected at the end of the semester then graded according to scoring rubrics by qualified research team members. The assessment process is illustrated in Figure 1.



Collected data was analyzed using descriptive and inferential statistics approaches. Proficiency rate, median, and mode for each item was calculated. Wilcoxon Signed-rank tests were conducted to examine the population proficiency on each item and task based on collected data (Sheskin, 2007).

### **INSTRUMENTATION**

Stiggins (1987) proposed a step-by-step guidance to develop valid performance-based assessment. The development process of the performance-based assessments used in the current study cooresponds with Stiggins' process. The sets of performance-based assessments were intentionally designed for the target student population in the Redesigned High Schools for Transformed STEM Learning Project, in order to challenge and evaluate the students on using cognitive abilities of application, analysis, evaluation, and creation in a scientific context at their grade level. These are the four higher-level cognitive abilities in Revised Bloom's taxonomy.

The assessment instrumentation was established based on North Carolina essential standards and Standard Course of Study Blueprint (NCNS, n.d.). The entire assessment instrument addressed five subjects: Biology, Chemistry, Earth/Environmental Science, Physical Science, and Physics. For each subject, a set of four or five semi-structured performance assessment tasks were created to guide student activities. Each competency-centered task included one or multiple activities with rubrics, addressing corresponding learning objectives (see Appendices A and B for sample assessments). Each activity consisted of several sub-activities, assessing student's performance such as research and investigation, brainstorming, exploration, and reflection. Accompanying assessment rubrics were established to score student artifacts.

The instruments mapped to the cognitive dimension of Revised Bloom's Taxonomy, and all activities reached higher-order skill requirements (Anderson et al. 2001). The interrater reliability of the instrument has been confirmed by a prior pilot study (Ernst & Glennie, 2015). The current study adopted Earth/Environment Science and Biology assessments. The five performance tasks implemented in the Earth/Environmental Science classes include:

- 1) Astronomy - Explain how the Earth's rotation and revolution about the Sun affect its shape and is related to seasons and tides.
- 2) Soil and Water Connections (Erosion) - Evaluate human influences on water quality in North Carolina's river basins, wetlands and tidal environments,
- 3) Clouds and Weather - Predict the weather using available weather maps and data (including surface, upper atmospheric winds, and satellite imagery).
- 4) Acid Rain - Analyze the impacts that human activities have on global climate change (such as burning hydrocarbons, greenhouse effect, and deforestation).
- 5) Build Your Own Seismometer (Earthquake) – Explain the probability of and preparation for geohazards such as landslides, avalanches, earthquakes and volcanoes in a particular area based on available data. (NCDPI, 2012c)

The five performance tasks implemented in the Biology classes include:

- 1) Inheritance and Molecular Genetics - Predict offspring ratios based on a variety of inheritance patterns (including dominance, co-dominance, incomplete dominance, multiple alleles, and sex-linked traits).
- 2) Cell Organelles - Summarize the structure and function of organelles in eukaryotic cells (including the nucleus, plasma membrane, cell wall, mitochondria, vacuoles, chloroplasts, and ribosomes) and ways that these organelles interact with each other to perform the function of the cell.
- 3) Photosynthesis and Cellular Respiration - Analyze photosynthesis and cellular respiration in terms of how energy is stored, released, and transferred within and between these systems.
- 4) Human Activities and the Environment – Infer how human activities (including population growth, pollution, global warming, burning of fossil fuels, habitat destruction and introduction of nonnative species) may impact the environment.
- 5) Enzyme and Fruit Browning - Explain how enzymes act as catalysts for biological reactions. (NCDPI, 2012b)

## FINDINGS

Student performances of each sub-activity were assessed based on established rubrics. Data from 63 student participants were collected regarding 71 assessed items (sub-activities), totaling 1088 scoring instances. An ordinal scale, from one to four, was adopted. The score of 1 represents “beginning to attain standard”; the score of 2 represents “nearly attained standard”; the score of 3 represents “achieved standards”; the score of 4 represents “exceeded standard”. Scoring 3 and above indicated that student’s performance reached proficient levels required by the state standards.

Descriptive statistics for the 10 performance tasks are summarized in Tables 3-12. Table 3 represents three performance activities and the measurable items within each activity. The first performance task was an astronomy task. This task included three activities:

1. Earth’s rotation
  - 1.1 required depiction using diagram
  - 1.2 required documentation and application
  - 1.3 required observation and depiction
  - 1.4 required explanation
2. Earth’s shape
  - 2.1 was an application task
  - 2.2 required observation, comparison, and explanation
  - 2.3 included application and explanation
3. Earth’s revolution
  - 3.1 was a diagramming task
  - 3.2 was a simulation and explanatory task

The astronomy assessment had a 46 percent proficiency rate among participants, indicating that 46 percent of the student artifacts reach proficient levels (scoring 3 or 4). The proficiency rates for each sub-activity ranged from 9 to 66 percent. Only 9 percent of students were proficient in sub-activity 2.3. Fewer than half of the students were proficient in sub-activity 1.1, 1.2, 1.4, and 2.2. Item 3.1, assessing student understanding through illustrating a scientific concept, had the highest median and mode. Table 3 summarizes the cognitive processes and categories based on Revised Bloom’s taxonomy, along with proficiency rate, median, and mode for each sub-activity in the astronomy task.

**Table 3.** Earth/Environmental Science - Astronomy Task

Item	Cognitive Process	Category	Proficiency Rate	Median	Mode
1.1	Illustrating	Understand	0.35	2	2
1.2	Implementing	Apply	0.48	2	1
1.3	Illustrating	Understand	0.64	3	3
1.4	Explaining	Understand	0.39	2	2
2.1	Implementing	Understand	0.66	3	4
2.2	Comparing and explaining	Understand	0.34	2	2
2.3	Implementing and explaining	Apply and understand	0.09	2	2
3.1	Illustrating	Understand	0.63	3.5	4
3.2	Implementing and explaining	Apply and understand	0.65	3	3

The second assessment task was the erosion performance task addressing soil and water connections. This task included one activity with six sub-activities:

- 4.1 required depiction with drawing a map
- 4.2 required research on consequences and relationships
- 4.3 required research on examples and impacts
- 4.4 required documentation of observations
- 4.5 required documentation of observations
- 4.6 required problem identification and proposing solution



The erosion activity performance-based outcomes produced a 49 percent proficiency rate. The proficiency rates for items ranged from 14 percent to 67 percent (see Table 4 for results on each sub-activity). Sub-activity 4.2 had the lowest proficiency rate (14 percent) and related to explaining the consequences through a cause-and-effect model.

**Table 4.** Earth/Environmental Science - Erosion Task

Item	Cognitive Process	Category	Proficiency Rate	Median	Mode
4.1	Illustrating	Understand	0.67	4	4
4.2	Explaining	Understand	0.14	2	2
4.3	Exemplifying	Understand	0.48	2	2
4.4	Representing	Understand	0.54	3	2
4.5	Representing	Understand	0.62	3	3
4.6	Designing	Create	0.42	2	3

The third performance task in Earth/Environmental Science was the cloud and weather task with seven sub-task assessment items:

- 5.1 required observation and documentation
- 5.2 required diagramming
- 5.3 required identification of indicators
- 5.4 required recording and analysis
- 5.5 required observation and prediction
- 5.6 required comparison
- 5.7 was an explanatory and reasoning task

The cloud and weather performance scores featured a 55 percent proficiency rate among student participants. All students who completed sub-activity 5.5 were proficient, but none of the students showed proficiency on sub-activity 5.6. Table 5 summarizes the cognitive requirements and student proficiency results for each sub-activity.

**Table 5.** Earth/Environmental Science - Cloud and Weather Task

Item	Cognitive Process	Category	Proficiency Rate	Median	Mode
5.1	Representing	Understand	0.5	3	4
5.2	Illustrating	Understand	0.8	4	4
5.3	Identifying	Remember	0.5	2	1
5.4	Organizing	Analyze	0.75	3	3
5.5	Predicting	Understand	1	3	3
5.6	Comparing	Understand	0	1.5	1
5.7	Inferring and explaining	Understand	0.25	2	2

The fourth task was the acid rain performance task with one activity consisting of six sub-activities:

- 6.1 required research and explanation
- 6.2 required observation and application
- 6.3 required observation and application
- 6.4 involved comparison
- 6.5 consisted of research summary and explanation
- 6.6 required solution proposal

The acid rain activity scores identified a 37 percent proficiency rate for student participants. Students had a low proficiency rate of around 15 percent on sub-activities 6.3, 6.5, and 6.6. The results suggested that participants tended to demonstrate lower proficiency on higher-order thinking skills than on understanding (see Table 6).

Table 6. Earth/Environmental Science - Acid Rain Task

Item	Cognitive Process	Category	Proficiency Rate	Median	Mode
6.1	Explaining	Understand	0.69	3	4
6.2	Implementing	Apply	0.33	2	2
6.3	Implementing	Apply	0.14	2	2
6.4	Comparing	Understand	0.69	3	3
6.5	Summarizing and explaining	Understand	0.15	2	2
6.6	Designing	Create	0.15	1	1

The earthquake task outcomes identified a 35 percent proficiency rate (see Table 7 for results on each sub-activity). This task involved a simulated experiment guided by six sub-activities:

- 7.1 was a research task
- 7.2 required estimation and documentation of experiment results
- 7.3 involved explanation
- 7.4 required interpretation
- 7.5 required shortcoming identification
- 7.6 required proposing solutions

Table 7. Earth/Environmental Science - Earthquake Task

Item	Cognitive Process	Category	Proficiency Rate	Median	Mode
7.1	Interpreting	Understand	0.27	2	2
7.2	Implementing	Apply	0.42	2	3
7.3	Explaining	Understand	0.26	2	2
7.4	Interpreting	Understand	0.13	1	1
7.5	Detecting	Evaluate	0.48	2	2
7.6	Designing	Create	0.55	3	2

Five performance tasks addressed biology subject matter. The inheritance task consisted of three activities regarding gene and inherit traits.

- 8. Human facial traits
  - 8.1 involved research and selection on facial traits
  - 8.2 required observation and documentation
  - 8.3 required application and analysis
  - 8.4 involved explanation and verification of a scientific concept
- 9. Human polygenic traits
  - 9.1 was a research task
  - 9.2 required classification
  - 9.3 required observation and documentation
  - 9.4 was an analysis task
- 10. Queen Victoria's hidden gene (recessive gene and family inheritance)
  - 10.1 required drawing conclusion
  - 10.2 required research and diagramming
  - 10.3 included proposing a solution

The performance scores featured a 47 percent proficiency rate among student participants based on the outcomes of the three activities. These sub-activities generally required understanding and applying scientific concepts to explain phenomena (see Table 8 for results on each sub-activity). Students had low proficiency rates on 8.3, 8.4, and 9.4, which targeted higher-order thinking skills. Sub-activities 10.1, 10.2, and 10.3, did not meet reporting requirements.



Table 8. Biology – Inheritance Task

Item	Cognitive Process	Category	Proficiency Rate	Median	Mode
8.1	Exemplifying	Understand	0.79	3	3
8.2	Exemplifying	Understand	0.69	4	4
8.3	Executing and organizing	Apply and analysis	0.08	2	2
8.4	Predicting and testing	Understand and evaluate	0.15	1	1
9.1	Exemplifying	Understand	1	3	3
9.2	Classifying	Understand	0.33	2	2
9.3	Identifying	Remember	0.5	2.5	4
9.4	Organizing and illustrating	Analysis and understand	0.08	2	2

Two activities composed the cell organelles performance task.

11. Cell organelle simile
  - 11.1 required to make simile
  - 11.2 required demonstration of conceptual understanding
  - 11.3 was an analysis task
12. Cell organelle concept map
  - 12.1 required research and depiction
  - 12.2 required depiction and analysis
  - 12.3 required a concept map

The cell organelles task had an overall 50 percent proficiency rate (see Table 9 for results on each sub-activity). Students had a zero proficiency rate on 12.2 and 12.3 because many of them did not address all twelve cell organelles required by the task and associated standard for reasons undetermined. This could be due to limited time or these performance tasks not being mandatory.

Table 9. Biology – Cell Organelles Task

Item	Cognitive Process	Category	Proficiency Rate	Median	Mode
11.1	Matching	Understand	1	3	3
11.2	Interpreting	Understand	0.5	2.5	2,3
11.3	Organizing	Analysis	1	3	3
12.1	Interpreting	Understand	0.5	2.5	2,3
12.2	Interpreting and organizing	Understand and analysis	0	2	2
12.3	Illustrating and organizing	Understand and analysis	0	2	2

The photosynthesis and cellular respiration task consisted of six sub-activities.

- 13.1 included research and diagramming on Photosynthesis
- 13.2 included research and diagramming on Respiration
- 13.3 required analysis of experimental results
- 13.4 involved experimental design
- 13.5 involved explanation
- 13.6 required experiment and documentation of findings

The proficient rate for this assessment was only 9 percent (see Table 10 for results on each sub-activity). The proficiency rate was zero on 13.1 and 13.3, and 14 to 17 percent on the other four sub-activities. Students had low performance on both understanding and higher-order skills for this task.

**Table 10.** Biology – Photosynthesis and Cellular Respiration Task

Item	Cognitive Process	Category	Proficiency Rate	Median	Mode
13.1	Illustrating	Understand	0	2	2
13.2	Illustrating	Understand	0.17	2	2
13.3	Differentiating	Analysis	0	2	2
13.4	Designing	Create	0.14	2	2
13.5	Explaining	Understand	0.14	1	1
13.6	Implementing	Apply	0.14	1	1

The Human activities and environment task consisted of two activities. Activity 14 focused on population, related policies and impacts, and consisted of two research sub-activities. Activity 15 concentrated on exploration and reflection on possible nearby pollution.

- 14. Impact of population
  - 14.1 required research and explanation
  - 14.2 required analysis on policies and their effects
- 15. Possible nearby pollution
  - 15.1 required observation and diagraming
  - 15.2 required depiction
  - 15.3 required research
  - 15.4 required analysis
  - 15.5 required proposing solutions

The proficient rate for this assessment was 33 percent (see Table 11 for results on each sub-activity).

**Table 11.** Biology – Human Activities Task

Item	Cognitive Process	Category	Proficiency Rate	Median	Mode
14.1	Explaining	Understand	0.4	2	2
14.2	Organizing	Analysis	0	2	2
15.1	Illustrating	Understand	0.45	2	1
15.2	Interpreting	Understand	0.55	3	3
15.3	Interpreting	Understand	0.18	1	1
15.4	Organizing	Analysis	0.45	2	3
15.5	Designing	Create	0.36	2	2

The enzyme task was built around enzymes functions and relevant biological reactions.

- 16.1 required research and explanation
- 16.2 involved research and factor identification
- 16.3 encompassed tabulating
- 16.4 involved observation and documentation of experiment
- 16.5 required data analysis skills
- 16.6 was an analysis task

Students had an overall proficiency rate of 67 percent for this assessment (see Table 12 for results on each sub-activity).

Table 12. Biology – Enzyme Task

Item	Cognitive Process	Category	Proficiency Rate	Median	Mode
16.1	Explaining	Understand	1	3	3
16.2	Identifying	Remember	0.81	3	3
16.3	Representing	Understand	0.88	4	4
16.4	Implementing	Apply	0.88	3	3
16.5	Organizing	Analysis	0.19	2	2
16.6	Differentiating	Analysis	0.25	2	2

Collective scores for each sub-activity have been examined to determine student performance on each sub-activity. In order to test the performance proficiency (scored three and above), the medians of collected ordinal data were compared to the cut-off value (specified parameter  $\geq 2.99$ ) using the nonparametric Wilcoxon Signed-rank test (Sheskin, 2007). Seventy-one Wilcoxon Signed-rank tests were implemented independently. Student performance for seven items were identified as proficient at the significance level of 0.05. The required performance and associated cognitive category of these items as well as the statistical outputs are shown in Table 13.

Table 13. Identified Performance Proficiency on Sub-activity Level

Item	Required Performance	Cognitive Category	Z-Score	p-value
4.1	Map drawing	Understand	137.5	0.0011
6.1	Research and explanation	Understand	35	0.0410
8.1	Research	Understand	28.5	0.0383
8.2	Observation and documentation	Understand	39	0.0002
16.1	Research and explanation	Understand	68	<0.0001
16.3	Tabulation	Understand	37	0.0199
16.4	Observation and documentation	Apply	44	0.0119

The collective data for each of the ten performance tasks were also tested via the Wilcoxon Signed-rank test. Only two, the cloud and weather task and the enzyme task, of the ten performance tasks were determined to meet the criteria (see Table 14). The number of instances involved in the statistical tests vary dependent on the number of constructs within each outcome variable.

Table 14. Identified Performance Proficiency on Activity (Task) Level

Task	Subject	Z-Score	p-value
Cloud and weather	Earth/ Environmental Science	280.5	< .0001
Enzyme and fruit browning	Biology	2328	< .0001

In addition, performance proficiency rates were tabulated by implementation site (see Table 15). Of the 63 student participants, Site 1 and 3 had the highest performance-based proficiency rate, followed by Site 4 and Site 5. However, at Site 2, less than one-fifth of the students were rated as proficient. This result did not show strong conformity to school performance on standardized tests (see Table 2 for school performance on end-of-course exams).

**Table 15.** Performance Proficiency by Site

Site	1	2	3	4	5
Subject	Earth/ Environmental Science	Earth/ Environmental Science	Earth/ Environmental Science	Biology	Biology
Proficiency Rate	0.505	0.179	0.505	0.439	0.372
Median	3	2	3	2	2
Mode	2	2	2	2	2
Range	3	3	3	3	3

### CONCLUSIONS AND RECOMMENDATIONS

School wide performance-based educational models have become prevalent, especially within STEM academy, magnet, and strand school formats (Ernst & Glennie, 2015). The redesigned STEM school model presents the opportunity to build performance task experiences into the academic learning environment. School-based constructivist approaches cultivate students with critical skills through the active participation in learning activities (Gulbahar & Tinmaz, 2006). These performance tasks provided students opportunities to interact with science concepts and applications, and observe and implement scientific investigation.

In the current study, students sporadically demonstrated higher-order proficiency. Students demonstrated proficiency specific to brainstorming through drawing maps, exploration through collecting and tabulating data, and research and investigation. Developing levels of performance-based application and scientific investigation skills have been shown regarding several indicators in some contexts (e.g. enzyme and fruit browning), however, the uniform proficiency has not been demonstrated. The results suggested that sampled students in general had not achieved the proficiency level of higher-order thinking skills in Biology and Earth/Environmental science.

The proficiency of certain indicators surfaced within some contexts but not in others. This failure of knowledge transfer might be related to the science content mastery. However, to diagnose the actual cognitive ability development progress, other types of assessments focusing on factual knowledge and conceptual understanding should be employed to accompany the results.

Students were sampled from ninth, tenth, and eleventh grade. Students who were in the earlier stages of secondary education might not have experienced performance-based learning to the same extent as students of advanced learner levels. In addition, a considerable portion of the student population in each participating school is identified as an underrepresented minority and/or low familial economic status (enrolling in Free- or Reduced-Price Lunch Programs). These students were conventionally underrepresented in STEM education and related careers. Students may need time to acclimate and/or develop knowledge and abilities gradually, in order to demonstrate desired proficiency on performance-based assessments.

Performance-based tasks and assessments allow the measuring on student procedural knowledge and higher-order thinking abilities, thus monitoring the teaching and learning outcomes under the new expectations of the current state standards. This research has potential to inform the NCNS curriculum and school model, as well as teacher classroom practice. The proficiency separations among sites might also indicate school climate, individual teachers, and their classroom practice impacts on student performance-based proficiency. Teacher's attitude and willingness of implementation as well as their strategies in scaffolding and supporting students might explain some variation. Research shows that teachers might consider that problem-based assessments were time-consuming and difficult to implement, and they might have not fully internalized the set methods and philosophy (Dağhan & Akkoyunlu, 2014). Future research that investigates the underlying causal mechanisms might contribute to the educational model and further development of the performance-based assessment techniques.

### AUTHOR BIOGRAPHIES

**Jeremy V. Ernst** is an Associate Professor in the School of Education at Virginia Tech. Jeremy specializes in research focused on dynamic intervention means for STEM education students categorized as at-risk of dropping out

of school. He also has curriculum research and development experiences in technology, engineering, and design education. Email: jvernst@vt.edu

**Elizabeth Glennie** is a Senior Research Analyst at RTI International. Her work focuses on the implementation and impact of educational policies on students, particularly the factors influencing success in secondary school and access to postsecondary education. She received her Ph.D. in Sociology from Duke University. Email: eglennie@rti.org

**Songze Li** is a Ph.D. candidate in the School of Education at Virginia Tech. Songze has recently served as a Graduate Research Assistant for the Re-designed High Schools for Transformed STEM Learning project. Her dissertation work is in STEM educator preparation to serve students with categorical disabilities or limited English proficiency. Email: lsz@vt.edu

## REFERENCES

- Alsardary, S., Pontiggia, L., Hamid, M., & Blumberg, P. (2011). Primary trait analysis to assess a learner-centered, upper-level mathematics course. *Primus: Problems, Resources & Issues in Mathematics Undergraduate Studies*, 21(4), 303-315. doi:10.1080/10511970903043755
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J. , & Wittock, M. C. (Eds.) (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York, NY: Longman.
- Arter, J. (1998, April). *Teaching about performance assessment*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Cydis, S. (2015). Authentic instruction and technology literacy. *Journal of Learning Design*, 8(1), 68-78.
- Dağhan, G., & Akkoyunlu, B. (2014). A qualitative study about performance based assessment methods used in information technologies lesson. *Educational Sciences: Theory & Practice*, 14(1), 333-338. doi:10.12738/estp.2014.1.2005
- Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64(1), 5-31. doi: 10.17763/haer.64.1.j57n353226536276
- Ernst, J.V. & Glennie, E. (2015). Redesigned high schools for transformed STEM learning: Performance assessment pilot outcome. *Journal of STEM Education*, 16(4).
- Falk, B., Wichterle Ort, S., & Moirs, K. (2007). Keeping the focus on the child: Supporting and reporting on teaching and learning with a classroom-based performance assessment system. *Educational Assessment*, 12(1), 47-75. doi:10.1207/s15326977ea1201\_3
- Gulbahar, Y., & Tinmaz, H. (2006). Implementing project-based learning and e-portfolio assessment in an undergraduate course. *Journal of Research on Technology in Education*, 38(3), 309-327.
- Hambleton, R. K., & Murphy, E. (1992). A psychometric perspective on authentic measurement. *Applied Measurement in Education*, 5(1), 1-16.
- Hibbard, K. M. (1996). *A teacher's guide to performance-based learning and assessment*. Alexandria VA: Association for Supervision and Curriculum Development.
- Kan, A. (2007). An alternative method in the new educational program from the point of performance-based assessment: Rubric scoring scales. *Educational Sciences: Theory & Practice*, 7(1), 144-152.
- McLoughlin, C., & Taji, A. (2005). Learner-centered approaches in the sciences. In C. McLoughlin & A. Taji (Eds.), *Teaching in the science: Learner-centered approaches* (pp. 1-8). New York, NY: Food Products Press.
- Morton, J., Cumming, A., & Cameron, H. (2007). Performance-based assessment in undergraduate medical education. *Clinical Teacher*, 4(1), 36-41. doi:10.1111/j.1743-498X.2007.00138.x
- National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Nitko, A. J., & Brookhart, S. M. (2015). *Educational assessment of students* (7th ed.). Boston: Pearson.
- North Carolina Department of Public Instruction. (2012a). High School Graduation Requirements. Raleigh, NC: Author.
- North Carolina Department of Public Instruction. (2012b). North Carolina Essential Standards: Biology. Raleigh, NC: Author.
- North Carolina Department of Public Instruction. (2012c). North Carolina Essential Standards: Earth/Environmental Science. Raleigh, NC: Author.
- North Carolina New Schools (2014). *Common Instructional Framework*. Accessed from <http://ncnewschools.org/strategy/tools-for-teaching-learning-and-leading/common-instructional-framework>
- North Carolina New Schools (n.d.). *NC essential standards*. Retrieved from <http://www.ncpublicschools.org/acre/standards/new-standards/>

- Oberg, C. (2009). Guiding classroom instruction through performance assessment. *Journal of Aviation Management & Education, 1*, 1-11.
- Pinter, B., Matchock, R. L., Charles, E. P., & Balch, W. R. (2014). A cross-sectional evaluation of student achievement using standardized and performance-based tests. *Teaching of Psychology, 41*(1), 20-27. doi: 10.1177/0098628313514174
- Rudner, L. M., & Boston, C. (1994). Performance assessment. *ERIC Review, 3*(1), 2-12.
- Sheskin, D. J. (2007). *Handbook of parametric and nonparametric statistical procedures* (4th ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Stiggins, R. J. (1987). Design and development of performance assessments. *Educational measurement: Issues and Practice, 6*(3), 33-42.
- Tanilon, J., Segers, M., Vedder, P., & Tillema, H. (2009). Development and validation of an admission test designed to assess samples of performance on academic tasks. *Studies in Educational Evaluation, 35*(4), 168-173.

**APPENDIX A**

**Sample Earth/Environment Science Assessment – Clouds and Weather Task**

Earth/ Environmental Science Blueprint 2.5.4:

Predict the weather using available weather maps and data (including surface, upper atmospheric winds, and satellite imagery).

Overview:

You are going to research and observe cloud patterns in order to analyze their connections with weather trends.

Phase I: Research and investigation (*approximately 45 minutes*)

1. Research different types of clouds; record in your notebook how they are formed, an estimate of their altitudes, and what upcoming weather they can possibly predict.
2. In your notebook, draw a picture for each type of cloud and explain how you can distinguish each cloud from others based on its appearance.
3. Research and record other observable weather indicators with explanation (for example, animal behavior, or wind direction) in your notebook.

Phase II: Exploration (*approximately 90 minutes*)

1. Select one of your identified weather indicators from Phase I, step 3. In your notebook, record your observation and analyze how it can help you predict weather.
2. Go outside of your home or school. Identify the cloud types; draw a picture of the clouds and record the time and date of observation.
3. Predict the upcoming weather based on your selected weather indicators and cloud observation.
4. Record weather patterns throughout the day. Make accurate notes including the start and duration times of precipitation as well as fair weather.
5. Perform steps 1 through step 4 for five days.

Phase III: Reflection (*approximately 20 minutes*)

1. Clip or print the weather forecast from newspaper or other media on the same day of your cloud observation; compare and analyze the forecast with your personal prediction.
2. Are contrails a type of cloud? Is fog a type of cloud? Research these questions, identify your answers, and provide explanations in your notebook.



## Earth/ Environmental Science 2.5.4 Assessment Rubric:

Scale	1 Beginning to Attain Standard	2 Nearly Attained Standard	3 Achieved Standard	4 Exceeded Standard
Cloud type research (Phase I, 1)	Cloud types were incorrectly recorded.	More than two cloud types were correctly recorded without all 3 required descriptions for each type.	More than four cloud types were correctly recorded with all 3 required descriptions for each type.	More than four cloud types were correctly recorded with all 3 required descriptions for each type. A detailed and logical explanation accompanies each description.
Drawing and cloud appearance (Phase I, 2)	Pictures were excluded or not related to cloud types.	Pictures and descriptions for all listed cloud types were not provided.	Pictures were drawn with eminent characteristics for all listed cloud types with descriptions for each cloud appearance.	Pictures were drawn with eminent characteristics for all listed cloud types with clear descriptions and logical explanations for each cloud appearance.
Weather indicators identification (Phase I, 3)	Weather indicators were excluded or could not be used to indicate weather.	One reasonable weather indicator was listed.	Two reasonable weather indicators were listed with logical description.	More than two reasonable weather indicators were listed and expanded upon with logical explanations of each indicator.
Observation recording and analysis (Phase II, 1,2)	Observation recorded was not related to the weather indicator. Picture drawn was not related to clouds.	Observation from selected weather indicator was recorded. Picture of cloud was drawn.	Observation from selected weather indicator was recorded and analyzed. Picture of cloud was drawn, cloud type was identified, and time and date documented.	Observation from selected weather indicator was recorded and analyzed. Picture of cloud was drawn, cloud type was identified, and an explanation of the cloud identification method was provided.
Weather prediction (Phase II, 3)	Prediction of weather was not documented.	Weather predicted was documented but not plausible.	Weather predicted was plausible based on selected weather indicator and cloud observation.	Weather predicted was plausible based on selected weather indicator and cloud observation; an explanation was provided for the weather prediction.
Prediction comparison (Phase III, 1)	Comparison was not made between student prediction and media prediction.	Comparison was made between student prediction and media prediction with an implausible analysis.	Comparison was made between student prediction and media prediction with a plausible analysis.	Comparison was made between student prediction and media prediction with plausible analysis. An explanation of the comparison was provided.
Fog and contrail (Phase III, 2)	Reasoning related to fog or contrail was not provided.	Reasoning for fog and contrail was implausible.	Reasoning for fog and contrail was plausible.	Reasoning for fog and contrail was plausible based on documented research.

**APPENDIX B**

**Sample Biology Assessment – Inheritance and Molecular Genetics Task**

Biology Blueprint 3.2.2:

Predict offspring ratios based on a variety of inheritance patterns (including dominance, co-dominance, incomplete dominance, multiple alleles, and sex-linked traits).

Activity 1: Human facial traits (*approximately 60 minutes*)

1. Research human physical facial traits that exhibit dominant or recessive characteristics. Select multiple traits on which you would like to conduct further experiments. Record your selections in your notebook.
2. According to your referred resources, write down the dominant and recessive characteristics for your selected facial traits in your notebook.
3. Look at the mirror and record your own facial traits in your notebook.
4. Look at your parents and record their facial traits in your notebook.
5. Draw Punnett squares for each trait to illustrate the connections between your parents and your facial traits in your notebook.
6. Prove the effectiveness of Mendel's Law of Independent Assortment.

Activity 2: Human Polygenic traits (*approximately 60 minutes*)

1. Research human traits that are of polygenic nature. Select one that you would like to use in conducting further experiments. Record your selection in your notebook.
2. Using your resources, write down the dominant, recessive, and intermediate characteristics along with the corresponding allele combinations in your notebook.
3. (After acquiring their permission) Collect information from five of your classmates, friends, or relatives regarding your selected characteristics in your notebook.
4. Collect information about both parents of your five selected classmates, friends, or relatives regarding your selected characteristics in your notebook.
5. Interpret your five collected sets of data using Punnett squares in your notebook.

Activity 3: Queen Victoria's hidden gene (*approximately 60 minutes*)

Queen Alexandrina Victoria was the monarch in Britain and Ireland in the 19<sup>th</sup> century. She had a recessive sex-linked X chromosome disorder—hemophilia, while her husband, Prince Albert, did not exhibit this disorder. Together they had nine children, four sons and five daughters.

1. Do you think that Prince Albert could also possibly carry recessive hemophilia disorder? Explain your ideas in your notebook.
2. Research on Queen Victoria's nine children, their spouses and Queen Victoria's grandchildren with regard to hemophilia disorder. Draw a pedigree diagram accordingly in your notebook.
3. According to your pedigree diagram, which marriages can serve the purpose of test cross of hemophilia disorder? Explain your answer in your notebook.

Biology 3.2.2 Assessment Rubrics

Activity 1:

<b>Scale</b>	<b>1 Beginning to Attain Standard</b>	<b>2 Nearly Attained Standard</b>	<b>3 Achieved Standard</b>	<b>4 Exceeded Standard</b>
Facial traits selection - dominant and recessive traits (step 1, 2)	Less than three facial traits were selected and dominant and recessive traits were incorrectly listed.	Less than three facial traits were selected or dominant and recessive traits were incorrectly listed.	Multiple facial traits were selected. Correct dominant and recessive traits were listed.	Multiple facial traits were selected. Correct dominant and recessive traits were listed. Some drawings were provided to aid visual conceptualization.
Data recording (step 3,4)	Data from the student or his/her parents were not recorded.	Less than three sets of data from the student or his/her parents were recorded.	Multiple sets of data from the student and his/her parents were recorded.	Multiple sets of data from the student and his/her parents were recorded in an orderly fashion.
Punnett square (step 5)	Punnett squares between parents were incorrectly drawn. Connections between the student and parents were not illustrated.	Punnett squares between parents were incorrectly drawn; or connections between the student and parents were illustrated incorrectly.	Punnett squares between parents for each trait were correctly drawn. Connections between the student and parents were illustrated correctly.	Punnett squares between parents for each trait were correctly drawn. Connections between the student and parents for each trait were illustrated correctly in a clear, logical, and detailed fashion.
Independent assortment (step 6)	Independent assortment was not proved.	Independent assortment was incorrectly proved.	Independent assortment was correctly proved.	Independent assortment was correctly proved. The human facial traits study was used as an example.

Activity 2:

<b>Scale</b>	<b>1 Beginning to Attain Standard</b>	<b>2 Nearly Attained Standard</b>	<b>3 Achieved Standard</b>	<b>4 Exceeded Standard</b>
Polygenic human traits selection (step 1)	Human trait was not selected.	The selected human trait was not of polygenic nature.	The selected human trait was of polygenic nature.	The selected human trait was of polygenic nature. Other polygenic human trait(s) was/were also listed.
Allele combination with dominant, recessive, and intermediate traits (step 2)	Alleles were incorrectly named. Dominant, recessive, and intermediate traits were not listed with corresponding allele combinations.	Alleles were correctly named. Dominant, recessive, and intermediate traits were partially listed with corresponding allele combinations.	Alleles were correctly named. Dominant, recessive, and intermediate traits were fully listed with corresponding allele combinations.	Alleles were correctly named. Dominant, recessive, and intermediate traits were fully listed with corresponding allele combinations. Some drawings were provided to aid visual conceptualization.
Data recording (step 3,4)	Data from the student's classmates, friends, or relatives and their parents were not recorded.	Less than five sets of data from the student's classmates, friends, or relatives and their parents were recorded.	Five sets of data from the student's classmates, friends, or relatives and their parents were recorded.	Five sets of data from the student's classmates, friends, or relatives and their parents were recorded in an orderly fashion.
Punnett square (step 5)	Punnett squares between parents were not drawn. Connections between the child and parents were not illustrated.	Punnett squares between parents were incorrectly drawn; or connections between the child and parents were illustrated incorrectly.	Punnett squares between parents were correctly drawn for all five sets. Connections between the child and parents were illustrated correctly.	Punnett squares between parents were correctly drawn for all five sets. Connections between the child and parents were illustrated correctly for all five sets in an organized and logical fashion.

Activity 3:

<b>Scale</b>	<b>1 Beginning to Attain Standard</b>	<b>2 Nearly Attained Standard</b>	<b>3 Achieved Standard</b>	<b>4 Exceeded Standard</b>
Hemophilia in men (step 1)	The student answered that Prince Albert could possibly carry recessive hemophilia disorder.	The student incorrectly explained that Prince Albert did not carry recessive hemophilia disorder.	The student correctly explained that Prince Albert did not carry recessive hemophilia disorder.	The student correctly explained that Prince Albert did not carry recessive hemophilia disorder in a logical and orderly fashion.
Pedigree (step 2)	The drawn hemophilia pedigree was incorrect with regard to Queen Victoria’s nine children, their spouses and Queen Victoria’s grandchildren.	The drawn hemophilia pedigree was partially correct with regard to Queen Victoria’s nine children, their spouses and Queen Victoria’s grandchildren.	The drawn hemophilia pedigree was fully correct with regard to Queen Victoria’s nine children, their spouses and Queen Victoria’s grandchildren.	The drawn hemophilia pedigree was fully correct with regard to Queen Victoria’s nine children, their spouses and Queen Victoria’s grandchildren in an orderly and easy-to-read fashion.
Test cross (step 3)	The student did not answer the questions.	The student incorrectly explained any marriage in which the husband did not exhibit hemophilia disorder.	The student correctly explained any marriage in which the husband did not exhibit hemophilia disorder. Understanding of test cross was shown.	The student correctly explained all marriages in which the husband did not exhibit hemophilia disorder. Adequate understanding of test cross was shown.