

This study was conducted by the Stanford Center for Opportunity Policy in Education (SCOPE) with support from the Ford Foundation and the Nellie Mae Education Foundation.

© 2010 Stanford Center for Opportunity Policy in Education. All rights reserved.

The Stanford Center for Opportunity Policy in Education (SCOPE) supports cross-disciplinary research, policy analysis, and practice that address issues of educational opportunity, access, equity, and diversity in the United States and internationally.

Citation: Abedi, J. (2010). *Performance assessments for English language learners*. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

Stanford Center for Opportunity Policy in Education

Barnum Center, 505 Lasuen Mall

Stanford, California 94305

Phone: 650.725.8600

scope@stanford.edu

<http://edpolicy.stanford.edu>



Table of Contents

Performance Assessments for English Language Learners	1
Why Performance Assessments for English Language Learners?	6
Improving Teaching Quality for English Learners Through Performance Assessments	12
How Performance Assessments Can Be Made Most Valid for English Language Learners.....	17
Summary and Discussion	21
References	23
Appendix: Sample PAT Response and Rubric	27

Abstract

Standardized achievement tests that are used for assessment and accountability purposes may not provide reliable and valid outcomes for English language learners (ELLs) because extraneous sources may confound the outcome of assessments for these students. Performance assessments, by contrast, may offer opportunities for these students to present a more comprehensive picture of what they know and are able to do. In this paper, we discuss limitations with the standardized achievement tests currently used for ELLs and share information on how performance assessments can be efficiently used to lead to better understanding of these students' content knowledge and to improve their academic performance. Issues and limitations of a performance assessment approach for ELL students are discussed, along with how to make these assessments more accessible to ELL students.

Acknowledgements

The author would like to acknowledge the contribution of Nancy Ewers, a graduate research assistant at the University of California, Davis, for her helpful editing, comments, and suggestions.

Performance Assessments for English Language Learners

Traditional standardized achievement test outcomes are used for high-stakes decisions in assessment and accountability systems throughout the United States. Often developed and field tested for the mainstream student population, these assessments may not be sensitive enough to the needs of some subgroups of students, such as English language learners (ELLs), who are faced with challenging academic careers. Research clearly demonstrates that some variables unrelated to the focal measurement construct (e.g., unnecessary linguistic complexity, cultural biases in construction of items) can affect the quality of high-stakes assessments for these students (Abedi, 2006; Solano-Flores & Li, 2006; Solano-Flores & Trumbull, 2003; Solano-Flores, 2008). Therefore, the outcomes of these assessments may not be reliable and valid, and they may not yield sufficient evidence for making important decisions regarding a student's academic career.

Despite efforts to make state and national standardized achievement tests more accessible for English language learners, the outcomes of these assessments may not be useful in evaluating student learning and informing instruction due to their inherent limitations. Conducted mainly for accountability purposes, these end-of-year assessments do not afford an opportunity for students to present a comprehensive picture of what they know and are able to do in content areas such as math, science, and reading/language arts. More importantly, “accountability is not only about measuring student learning, but actually improving it” (Darling-Hammond, 2004, p. 1078).

Performance assessments can help to fill this gap, because they not only engage these students and give them a chance to demonstrate their knowledge but also disclose more in-depth information on students' academic needs. Performance assessments can be less affected by unnecessary linguistic complexity for two reasons. First, language is often not the only medium of presenting an assessment task. For example, in a science “hands-on performance” task, students are presented with a set of physical materials (batteries, wires, and bulbs as in Figure 1 (see page 2), or pencil and salt and fresh water as in Figure 2, page 3). Second, students have access to these physical materials as they formulate assessment responses thereby reducing reliance on language. In science, students can experience manipulation of hands-on objects and use that familiarity as they formulate written or oral responses. Performance assessments thus reduce linguistic complexity as tasks are presented and as responses are prepared.

This report describes how performance assessments can help the ELL student population demonstrate what they know and are able to do. Unfortunately, research literature on performance assessments for English language learners is thin, but it offers evidence on the effectiveness and usefulness of performance assessments for these students.

Figure 1. Electric Mysteries Performance Assessment



“Electric Mysteries: Student determines what is inside an electric mystery box by constructing and reasoning about circuits. Scoring is evidenced based, focusing on evidence and explanation” (Ayala et al., 2001, p. 25).

Performance Assessments and English Language Learners

Bass, Magone, and Glaser (2002) observe that performance assessments allow all students, especially those with different language backgrounds, to engage in cognitively complex activities such as generating strategies, monitoring work, analyzing information, and applying reasoning skills.

At the same time, assessment experts may argue that for ELL students, performance assessment items may not fully reflect content knowledge and may be confounded with writing skill and vocabulary. That is, language factors in performance assessments may even have a greater level of impact on ELL student performance than for native speakers of English. In response to this concern, a distinction should be made between language related to the construct being measured (construct-relevant) and language not necessarily relevant to the content (construct-irrelevant). In performance assessments, students’ actual performance rather than their expressive language may more clearly convey the content being measured. For example, in the floating pencil exercise (see Figure 2, page 3, for a description), students not only hear the test instruction to measure the length of a pencil floating in both fresh and salt water but also see the pencil under both conditions. More importantly, since the object of measurement is presented in multiple sensory modes (floating pencil and fresh and salt water), language is not the only way to present the performance assessment task (PAT) and obtain student responses.

As Linn and Burton (1994, p. 5) have indicated, performance assessments have appeal as assessments that better reflect good instructional activities, are often thought to be more engaging for students, and are better reflections of criterion performances that are of importance outside the classroom (i.e., they are said to be more authentic).

Similarly, as Darling-Hammond (2006) indicates:

Performance assessments that require students to evaluate and solve complex problems, conduct research, write extensively, and demonstrate their learning in projects, papers, and exhibitions have proven key to motivating students and attaining high levels of learning in redesigned high schools (p. 655).

Increasing the level of motivation is important for English learners because they need encouragement and support in their academic endeavors.

Research clearly demonstrates that language factors have a major impact on the outcomes of assessments for English language learners (Solano-Flores & Li, 2006; Solano-Flores & Trumbull, 2003; Solano-Flores, 2008). Inclusion of unnecessary linguistic complexity within an assessment leads to a widening in the performance gap between ELL and non-ELL students. Although language factors influence any kind of assessment, performance assessments can actually help to identify language factors that influence assessment outcomes. For example, Mislevy, Steinberg, and Almond (2002) showed that task-based language assessments (TBLAs) can assess language in more realistic and complex settings than traditional discrete-skills assessments can, with the latter offering narrower, more artificial opportunities for receptive and expressive language use. Another recent study found student responses to a writing prompt less affected by student background variables, including English learner status, than were scores on a commercially developed language arts test, largely comprising multiple

Figure 2. A Hands-on Performance Item (Fourth Grade Task)

The fourth grade Floating Pencil task is intended to measure students' ability to collect data (measure length and volume), make inferences, and apply their understanding to new situations. In the task, students are told that they can determine the difference between fresh water and salt water by doing a test. First, students are instructed to measure the length of a pencil weighted [vertically] with a thumbtack (which serves as a hydrometer) [and] floating [partially] above the surface of the water, in both fresh water and salt water. The pencil is marked with equally spaced letters from A (top of pencil) through J (bottom of pencil), and students are asked to observe where the water line comes to on the pencil and place a mark on a picture of the pencil. Students are then directed to measure the length of the pencil that was above the water using a to-scale picture of a ruler. They repeat the Floating Pencil test to identify a "mystery water," measuring the length of pencil floating above the water in the mystery water and comparing this finding with results from the previous tests. Throughout the task, students also are asked (1) whether the amount of water in the cylinder changes when the pencil is added; (2) how the way the pencil floats in salt water compares with how it floats in fresh water; (3) how dissolving more salt in the salt water would change the way the pencil floats; (4) how they can tell what the mystery water is; and (5) whether, when people are swimming, it is easier for them to stay afloat in the ocean or in a fresh-water lake.

From Bass, Magone, & Glaser (2002), p. 5.

choice items (Goldschmidt, Martinez, Niemi, & Baker, 2007). Open-ended assessments improve the chances for ELL students to engage with language production and learning, offering unique opportunities for ELL students to express their knowledge in a broader sense than the limited linguistic opportunities given to them in traditional multiple choice items.

One goal of a performance assessment is to judge the level of competency students achieve in doing reading/language arts, science, and mathematics (Parker, Louie, & O'Dwyer, 2009). Therefore, performance assessments can also produce useful information for diagnostic purposes to assess what students know, and they can help teachers decide where to begin instruction or determine which groups of students need special attention. These assessment strategies can also be used to monitor students' processing skills and problem-solving approaches, as well as their competence in particular areas while simulating learning activities. These characteristics can be extremely beneficial for special needs student populations, including ELLs, since these students may not have received equal education opportunities because of their linguistic needs (Abedi & Herman, 2010). These students often exhibit greater interest and a higher level of learning when they are required to organize facts around major concepts and actively construct their own understanding of the concepts in a rich variety of contexts.

Performance tasks are also instructional, allowing students to actively engage in worthwhile learning activities within the classroom. In performance assessment settings, students may be encouraged to seek out additional information or try various approaches, and in some situations work in teams. These assessment strategies are all beneficial for ELLs as the students benefit from engagement in classroom activities. Furthermore, performance assessments are more accessible because many of the variables affecting large-scale state and national assessments have less impact on performance assessments and learning environments (Boscardin, Aguirre-Munoz, Chinen, Leon, & Shin, 2004; Wang, Niemi, & Wang, 2007a, 2007b). For example, on the basis of the findings of their study, Wang and colleagues (2007a) indicated that performance assessment outcomes are not sensitive to elements of students' background status such as SES and ethnicity. In other words, student strengths were more fully demonstrated on the performance assessments without undue influence of some of the sources of construct-irrelevant variables.

Performance assessments can be presented in many forms, yet are comprehensive in nature and allow students to present a more thorough indication of their understanding of certain content areas. A prime example of a performance assessment is a situation where students are asked to actively communicate in a second language or design and conduct research on a topic of interest. In this situation, the ELL students' speaking and writing abilities could be directly evaluated on the basis of the actual presentations and texts that are created by these students.

Linn, Baker, and Dunbar (1991) indicated that direct assessments of writing, for example, provide instances of the tasks that we would like students to be able to perform,

whereas questions about proper grammar are designed to measure what are best termed enabling skills or partial indicators of actual ability to write (p. 4).

Indirect proxies for the knowledge or skills to be measured can raise validity concerns. For example, assessing writing by asking questions about grammar may not be highly correlated with actual writing ability. Furthermore, questions about nuances of grammar and syntax may be particularly disadvantageous for ELL students whose first language is based on differing rules. By contrast, direct assessments of writing give students more opportunity to demonstrate their ability to convey ideas.

More examples appear throughout the report to illustrate how performance assessment tasks can help improve the quality of education for ELL students. Additionally, these examples compare and contrast the quality of measurement outcomes of performance assessments against those from traditional standardized achievement tests. This helps demonstrate that outcomes from performance assessments can be more informative for teachers, students, and parents of all students, particularly English language learners; these assessments are not as greatly affected by extraneous variables as are traditional high-stakes assessments (Boscardin et al., 2004; Wang et al., 2007a, 2007b).

Why Performance Assessments for English Language Learners?

Limitations of Standardized Assessments for ELLs

Standardized achievement tests are commonly used for assessment and accountability purposes because they have established objectivity, reliability, and validity as well as ease of administration and cost efficiency in scoring (Burger & Burger, 1994). These tests may also refer more succinctly to content standards and have more easily verifiable content representation than many current classroom measures (Chung, Delacruz, & Bewley, 2006; Mehrens, 1992). Research has pointed to many limitations in achievement tests used for student assessment and accountability purposes (Linn et al., 1991; Archbald & Newman, 1988; Shepard, 1991). These issues and limitations are particularly serious for ELLs who are often at the lower levels of academic performance distribution. Research shows that high-stakes testing policies can create inequity for low-achieving students when schools or districts systematically exclude them from these assessments in order to demonstrate gain in overall student achievement. For example, Heilig and Darling-Hammond (2008) found widespread exclusion of English learners from testing in a high-stakes accountability environment, both through mechanisms that rendered them “missing” on testing days and through policies and practices that pushed older students out of school entirely. Other studies on grade retention, exclusion from state and local assessment testing (and school), and drop-out rate also identify these sources of gaming.

These findings raise an important concern for English language learners, because they are often among the lowest-performing students in schools thanks to inequity issues in their instruction and assessments. If they are being categorically excluded from grade-appropriate high-stakes assessments in order to inflate a school’s or district’s scores, then their progression and achievement in school may be misrepresented in schools that rely solely on standardized tests as a barometer of performance.

At the same time, it may not help to include these students without appropriately assessing them. As Ayala, Shavelson, and Ayala (2001) indicated, “Although multiple-choice tests are useful for ascertaining a child’s conceptual knowledge, an assessment of actual performance may be more appropriate in some situations” (p. 1). Miller and Linn (2000) presented several reasons for using performance assessments instead of standardized tests. Among them is concern about the possible unintended negative effects of multiple-choice (MC) assessments, such as the belief that they lead to narrower curriculum and teaching to the test.

There are other technical issues with MC assessments. For example, a major threat to their validity is the technical problems associated with distractors in MC items. In addi-

tion to the possibility of unequal frequencies of response across the distractors, various subgroups of students may show dissimilar trends in selecting distractors when they are not sure about the correct response. For example, Abedi, Leon, and Kao (2008) found the pattern of selecting distractors to vary by students' disability status. Students without disabilities use the distractors that are more likely to be the correct answer, while students with disabilities randomly select distractors. Such differences in the pattern of selecting distractors may also seriously influence performance of ELL students. Abedi et al. (2008) found students at a lower level of proficiency (particularly those with disabilities) select distractors randomly rather than through educated guesses. Similar situations may apply to ELL students. If these students do not know the correct response to a multiple choice question that requires fine discrimination among terms, they may select a distractor that is simple and less linguistically challenging.

Another example of selective distractor patterns among ELLs was found in the responses of grade 8 students on multiple choice standardized assessments across three states (Abedi et al., 2010). Five items with the highest level of variation between distractors for ELLs were examined. ELL students most often selected distractors with a high incidence of academic vocabulary. It appeared that these distractors with academic terms were more attractive, and ELL students tended to select those responses in spite of being distractors. (The actual items could not be presented because of confidentiality issues.)

Issues of Linguistic Complexity

Analyses of national and state data show a substantial gap between the academic performances of ELLs contrasted with native speakers of English (Abedi, 2006, 2008). Abedi, Leon, and Mirocha (2003) compared the performance of ELL and non-ELLs across several content areas and states. Results of their analyses suggest that the higher the level of language demand of the test items, the larger the performance gap between ELLs and non-ELLs.

To illustrate this point, we present some of the findings of the analyses of NAEP data by Abedi et al. (2003). Results of these analyses show that the construct-relevant performance differences between ELLs and non-ELLs in grades 10 and 11 are highest in reading, a subject with a significant language demand that is considered "construct-relevant" since the focal construct is language. The performance gap decreases in science and mathematics. Averaging overall scores for students in grades 10 and 11, the performance gap between ELL and non-ELL students was 15.0 score points (on the Normal Curve Equivalence or NCE scale score) in reading, which was reduced to 10.5 points in science and further reduced to 1.3 points in math computation.

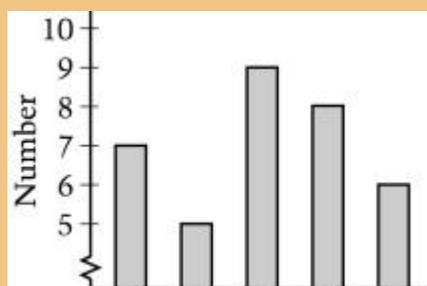
To further illustrate the impact of language factors on assessment outcomes for ELLs, Abedi (2008) analyzed data from several states. Comparisons were made between the

performance of ELLs and non-ELLs using a Disparity Index (DI), which is simply the percentage of difference between the two groups. Conducted on pre-NCLB (No Child Left Behind) data, these analyses again showed that the performance gap between ELLs and non-ELLs is higher in areas with more complex linguistic structure than in those with less complex linguistic demands. For example, the DI for ELLs in reading for grade 3 students was -53.4, suggesting that ELLs underperformed non-ELLs by 53.4%. For math, the DI for these students was 14.5%, substantially lower than those reported above for reading. Results from Site 3 furnish another example. Those results are consistent with results from Site 1; the data suggest a substantial performance gap between ELLs and non-ELLs. The Disparity Index (DI) for ELLs in reading for grade 5 students was -33.4, suggesting that ELLs performed 33.4% lower than non-ELLs as compared with a DI of -22.6 in math.

Abedi (2008) also conducted analyses on post-NCLB data and observed a trend quite similar to that of pre-NCLB data. These results again suggest that (1) ELLs perform substantially lower in all content areas than non-ELLs, and (2) the performance gap between ELL and non-ELL students is lower in math than in reading, implying that language factors play a major role in this gap.

Figure 3 (below)—a short item from one state’s released test items—includes passive voice, a complex verb, two relative pronouns, two noun phrases, and four different entities—all grammatical components that are contributors to complexity (Abedi et al., 2010).

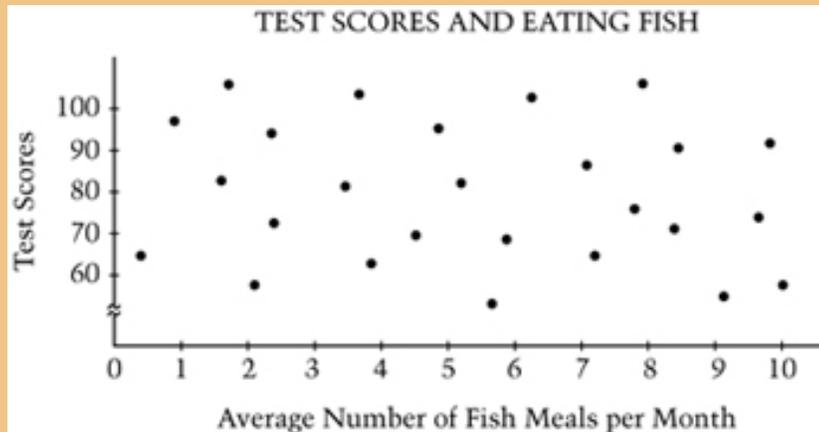
Figure 3. Example of a Linguistically Complex Multiple Choice Item



Jim made the graph above. Which of these could be the title for the graph?

- (A) Number of students who walked to school on Monday through Friday
- (B) Number of dogs in five states
- (C) Number of bottles collected by three students
- (D) Number of students in each of ten clubs

Figure 4. A NAEP Math Item for 8th Graders



For a science project, Marsha made the scatterplot above that gives the test scores for the students in her math class and the corresponding average number of fish meals per month. According to the scatterplot, what is the relationship between test scores and the average number of fish meals per month?

- (A) There appears to be no relationship.
- (B) Students who eat fish more often score higher on tests.
- (C) Students who eat fish more often score lower on tests.
- (D) Students who eat fish 4–6 times per month score higher on tests than those who do not eat fish that often.
- (E) Students who eat fish 7 times per month score lower on tests than those who do not eat fish that often.

Source: National Center for Education Statistics, 2005, grade 8, year 2007, item 13, block M7.

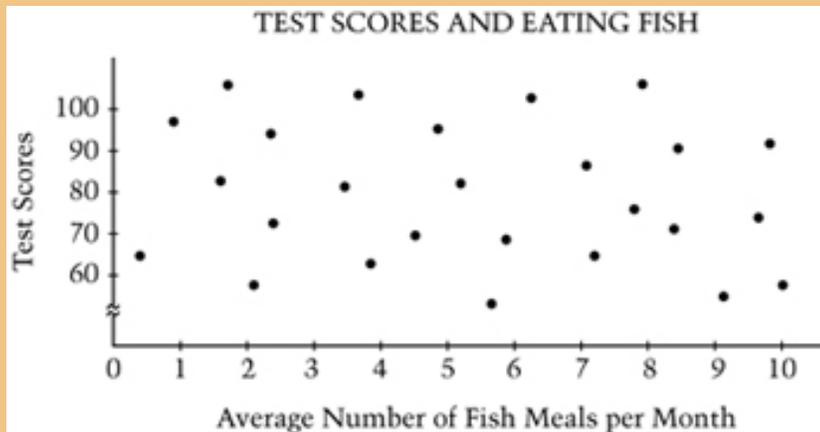
Figure 4 (above) is another example of a multiple choice assessment question that could be challenging for ELL students. It presents a math item from the 2007 National Assessment of Educational Progress (NAEP) for grade 8 students.

Although this item may work well for non-ELL students, it is less likely to function well for ELL students for two reasons. First, both the stem and choices for this item are long. As indicated earlier, Abedi, Lord, and Plummer (1997) found that ELL students performed substantially lower than their non-ELL peers on NAEP items with more than three lines of stem and more than a line on any of the choices. Second and more impor-

tantly, the choices for this item have varying lengths. The first three choices are relatively short and the last two are long. ELL students, particularly those with lower English proficiency, may select shorter choices just to be done with it, particularly if they have difficulty understanding the language in the long choices. This could create a larger performance gap between ELLs and non-ELLs, as well as affect the scoring and scaling of test items. Consequently, this may cause differential distractor selection patterns (see, for example, Abedi et al., 2008), which influence scoring of the items.

In addition to the item length and distractor issues, there are linguistic complexity issues in this item for ELLs. For instance, two experts rated the grammatical complexities of this item by six linguistic features: (1) passive voice, (2) complex verb, (3) relative clause, (4) subordinate clause, (5) noun phrase, and (6) entities. These

Figure 5. A Linguistically Modified NAEP Multiple Choice Item



The scatterplot above shows the relationship between students' math scores and number of hours reading a newspaper. On the basis of this scatterplot, what can you conclude?

- (A) There seems to be no relationship between math performance and number of hours reading the newspaper.
- (B) The more hours reading the newspaper the higher the math score.
- (C) The more hours reading the newspaper the lower the math score.
- (D) Students reading the newspaper 7 hours a week score lower on tests than those not reading the newspaper.

features were found to slow down the readers and make interpretation of text difficult (Abedi et al., 2010; Abedi & Lord, 2001). The two raters consistently identified eight instances of relative clauses, six instances of noun phrases, and six instances of entities. Furthermore, this item combines two activities that could easily be outside the cultural/linguistic understanding of ELLs: taking tests and eating fish. A more obvious non relationship is shown in Figure 5 (see page 10).

Performance assessment tasks may also suffer from unnecessary linguistic complexity because students have to read test directions and contextual information in the items. These may be linguistically complex readings. Students also have to write to explain and justify their responses, adding to the linguistic demands of these tasks. The content in which linguistic complexity occurs is a major distinction between the two types of assessments in terms of linguistic structure. In standardized test items, it is difficult to differentiate between complex linguistic structures that are the target of assessment and those that are unrelated to the construct being measured. It is also difficult to ascertain whether student responses result from linguistic challenges with the prompt or with the set of multiple-choice answers, including distractors. In performance assessment tasks, complexity may occur mostly in areas not related to the construct being measured (directions, context, etc.). Therefore it is less challenging to simplify the linguistic structure in PATs without altering the construct being measured. We next discuss how this can be done.

Improving Teaching Quality for English Learners Through Performance Assessments

Performance assessments make a major contribution to the academic careers of students, particularly those with challenging academic lives, by informing instruction and supporting higher-quality teaching and learning (Darling-Hammond, 2007). Miller and Linn (2000) believe that “much of the impetus for [performance assessments] is that they should mirror the teaching and learning process and provide a better measure of accountability” (p. 373). The authors reported on teacher attitudes and practices toward state-mandated performance assessments in five states. The results showed that teachers tried to align instruction with state performance assessments and supported using this type of mandatory assessment to improve instruction.

Firestone, Mayrowetz, and Fairman (1998) point out that “performance-based assessment can change specific behaviors and procedures in the classroom more easily than the general paradigm for teaching a subject” (p. 111). They suggest that working closely with performance assessments may improve teachers’ instructional knowledge. Similarly, Linn et al. (1991) indicated that teaching for test items emphasizing problem solving is not limited to a single right answer; rather there could be many ways of solving problems—an important aspect of good instruction.

Analyses of student writing reveal that students often lack understanding of expected language use in performing academic tasks (Aguirre-Munoz et al., 2006). In a study addressing this shortfall, Aguirre-Munoz et al. (2006) trained teachers using a comprehensive functional linguistic approach to academic language, which enhances a student’s own interpretation of characters and events (Christie, 1986, 2002). They found that teachers believed this approach has potential in improving ELL writing development. It offers more explicit instruction and permits performance assessment of genre-specific features of academic language to enhance reading comprehension and writing skills.

Using the outcomes from performance assessment to inform instruction could be of great value for English learners, because it helps their teachers improve the quality of instruction for these students. For example, Abedi and Herman (2010) found that ELL students report a lower level of opportunity to learn (OTL)—defined as content coverage—than their native English speaking peers, even within the same classroom. These findings may suggest that ELL students did not fully benefit from teachers’ instructions, possibly because of language issues.

Performance assessment outcomes can yield valuable information. For instance, one could set performance assessments to directly involve ELL students in classroom conversation and encourage them to report their understanding of teachers’ instructions, particularly with complex language instruction. The outcome would help teachers evaluate a student’s listening and speaking knowledge through observation and student presentations.

Assessment literature has introduced many forms of performance assessments for the general student population, including portfolio assessment, hands-on performance, and direct observation of student performance (Shavelson, Gao, & Baxter, 1993). Portfolio assessment is one of the most commonly used forms of performance assessment (Firestone et al., 1998). Firestone and colleagues believe that scores from portfolio assessments can furnish a comprehensive view of student work and teacher assignments. Linn et al. (1991) presented examples of direct assessments of complex performance. They include open-ended questions, essays, hands-on performance (see also Shavelson, Solano-Flores, & Ruiz-Primo, 1998; Shavelson, Baxter, & Pine, 1991), computer simulations or real-world problems, and portfolios of student work. Linn and colleagues indicated that “collectively such measures are frequently referred to as authentic assessments . . . because they involve the performance of tasks that are valued in their own right” (1991, p. 15).

In discussing design and analysis in task-based language assessment (TBLA), Mislevy et al. (2002) showed that linguistic competency, which includes knowledge of vocabulary and grammar, is not sufficient to evaluate communicative competence. The authors pointed out that “the concern in TBLA extends beyond knowledge and language per se, to the ability to deploy language knowledge appropriately and effectively in educationally or professionally important language use settings” (Mislevy et al., 2002, p. 3). The authors believe that an assessment such as TBLA, which includes sociolinguistic competence, strategic competence, and discourse competence, can present a broader conception of communicative competence. They refer to the listening portion of the Test of English as a Foreign Language (TOEFL) as an example of a TBLA in which all items in the test require information in the form of a short conversation in a college environment.

Boscardin et al. (2004) designed the Language Arts Performance Assessments (LAPA), requiring students to construct their own responses to open-ended prompts about literary works involving substantial integration of text-based information. They generate outcomes not easily obtained with multiple-choice or short constructed-response questions. Bass et al. (2002) used think-aloud protocol to examine the thinking and reasoning of fourth and eighth grade students engaged in two hands-on science tasks in NAEP. The think-aloud protocol would be valuable for ELL students because, through this approach, teachers and others involved in educating ELL students can obtain a clearer picture of where content and language can be confounded.

The common characteristic inherent in all ELL students is their need for linguistic support. Therefore, performance assessments supplying comprehensive information on the student’s level of English proficiency would be extremely helpful in informing curriculum and instructional planning. Performance assessment outcomes focusing on English language content would furnish information needed to understand how unnecessary linguistic complexity influences assessing English learners. Performance assess-

ments can enable students to demonstrate their knowledge unencumbered by confusion caused by linguistic complexity.

Unfortunately, research is scarce on performance assessments in English language content for ELL students. Given all the positive contributions that performance assessment outcomes can have on academic performance of these students, more effort is needed in developing and applying performance assessments in examining students' English language knowledge. This paper is the first step in such an endeavor.

Performance Assessment Approach in Testing Students' Content Knowledge in English

ELL students' knowledge of the English language is assessed to (1) determine their level of English proficiency as required by NCLB Title III and (2) measure their literacy level in English as required by NCLB Title I. Both measures influence ELL students' academic careers. We chose Title III English proficiency assessments to illustrate applying performance assessment procedures to assess English language competence for ELL students. The outcomes of these assessments are the most important because they determine students' ELL status and may influence instruction. NCLB Title III requires assessing ELL students' level of English proficiency in the four domains of reading, writing, listening, and speaking (NCLB, 2002). These assessments should be based on the concept of academic English and aligned with the states' English language proficiency and content standards.

Prior to NCLB, there were many tests for measuring ELL students' level of English proficiency. These assessments, which were mainly composed of multiple-choice questions, had dissimilar content, measured different language constructs, and had a number of formats and psychometric properties (Abedi, 2007, 2008; Parker et al., 2009). After NCLB, several consortia of states developed and field-tested batteries of ELP tests. Reviewers of these more recent assessments indicated areas that needed improvement (see, for example, Abedi, 2008). The listening and reading components of the tests use a traditional multiple-choice item format; the speaking and writing components are in the form of performance-based tasks and are scored against their respective rubrics. In the speaking test, tasks are administered individually in an interview format. Writing items are in a short-answer or essay format and administered in a group setting (Bauman, Boals, Cranley, Gottlieb, & Kenyon, 2007). The performance assessment format was clearly the optimal approach for these assessments.

We believe assessment of ELL students' English proficiency through performance assessment procedures offers comprehensive and useful outcomes. The next section illustrates how a performance assessment approach can be used in assessing students' English proficiency.

Performance Assessment: Illustrative Examples

This section presents two examples of performance assessment that have been developed by an educator with both teaching and assessment experience. They illustrate the format and use of performance assessments for ELL students. The first example focuses on applying performance assessment in reading and writing, and the second shows how to apply a performance assessment approach in social studies. The former is of particular relevance because language assessment is one of the most important aspects of ELL students' academic careers.

Case Example 1: Reading and Writing

Adele Fiderer is a fourth grade classroom teacher and language arts developer who has incorporated performance assessment into her teaching practice. In her reading class, Adele always encouraged students to choose their own books and read them independently, and then she would ask her students to talk and write about what they were reading. Throughout the years she developed portfolios of each student's best work, but she thought this was not a completely accurate measure of a student's learning. As Adele was looking for a comprehensive assessment technique to support the natural act of reading and responding to a story, she discovered through research that the performance assessment approach would give a better picture of what she wanted to know about her students' reading performance.

She adopted a performance assessment that required a common reading and writing task for the entire class. She begins by selecting a text that her students have not read—for example, a book discussing a story with a significant theme appropriate for her students that has a clearly identifiable problem and resolution, well-developed characters, and high interest. Adele then creates a writing task that encourages students to think about the story. Before they write their final draft, she gives them prewriting organizers such as webs, maps, and Venn diagrams, allowing her students to make notes about their ideas and giving them enough time to complete the performance assessment. Such a writing prompt for primary students would involve asking them to think about how to retell a story they have read to a friend. Students prewrite using a story map to outline all the important parts of the story and then write about the story on lined paper in final draft form.

Adele developed a rubric with ratings on a scale of 0–3 to evaluate children's writings about a story problem. Within the rubric, she provides detailed information of the performance level descriptions for all four levels (0–3). For example, to obtain a perfect score of 3, a student's written response must be complete, indicating good understanding of the story and its problem, and it must give accurate and relevant details, information, and supportive reasoning.

Case Example 2: Social Studies

Adele strongly believes that the process used to develop performance assessments in reading and writing can be used for other subject areas. For social studies, she begins the development process by looking for interesting nonfiction stories and articles relating directly to students' current studies in history and social studies, and then she designs a writing prompt asking them to think in depth about the subject. She cites an example of this process as it relates to the topic of immigration within her school's sixth grade curriculum:

[The] sixth graders read "The Letters of Rosie O'Brien, a Convict in New South Wales" (Cobblestone, April and May 1987). On one side of a planning sheet, students listed words and phrases that they felt described Rosie. On the other side of the sheet they provided evidence from the story that backed up their opinions. The students were then asked to use their knowledge of Rosie's character to write a letter from Rosie to her sister (Fiderer, 2009).

How Performance Assessments Can Be Made Most Valid for English Language Learners

Because the language of assessment is among the most influential factors affecting the outcomes of assessment for ELL students, we elaborate on the impact of such factors and offer recommendations on how to improve the quality of performance assessment with more linguistically accessible outcomes. As we noted, ELL students and students with disabilities sometimes perform better on performance tasks than on multiple-choice tests. This was the case in the New Jersey Special Review Assessments offered to students failing the state high school exit exam. These open-ended performance tasks test the same standards and concepts as items on the multiple-choice test but have proved more accessible to these populations of students.

Nonetheless, in any kind of test, careful design can make a difference in validity for special populations. As I mentioned earlier, research on ELL students has identified a number of linguistic features of test items that slow readers down and increase the chances of misinterpretation, among them language load, complex linguistic structures, and length (Abedi et al., 1997). Researchers have found that linguistic modifications reducing the complexity of sentence structures and replacing unfamiliar vocabulary with more familiar words increase the performance of English learners, as well as other students in low- and average-level classes (Abedi & Lord, 2001).

Linguistic modifications can be used in designing performance assessments to help ensure a valid and fair assessment, not only for English language learners but for other students having difficulty with reading. Figure 6 (page 18) shows how a task from the New Jersey SRA can be made even more accessible with linguistic modifications, without altering the knowledge and skills being measured. These modifications reduce the length of the task by more than 25% (from 264 words to 184), eliminate conditional clauses and grammatical complexities (such as passive voice), and use more familiar words. Although the modified task is easier to read, it still tests the same mathematics skills.

It is important to note that linguistically sound items do not avoid technical language appropriate to the content being measured. Figure 7 (page 19) shows a PAT in mathematics from the 2002–03 New Jersey SRA, High School Proficiency Assessment (HSPA). The language in this item, which includes sophisticated mathematical terms, is related to the content being measured and has a minimum level of unnecessary linguistic complexity.

Performance assessments presented in clear language and not affected by cultural biases may offer a better opportunity for ELLs to present a valid picture of what they know and are able to do. We believe that using a performance assessment approach to measure ELLs' content knowledge has the potential of incorporating research findings on how to reduce sources of threats to accessibility of assessment for these students.

Figure 6. New Jersey Department of Education, 2002–03 SRA Mathematics Performance Assessment Task

ORIGINAL ITEM: Dorothy is running for president of the student body and wants to create campaign posters to hang throughout the school. She has determined that there are four main hallways that need six posters each. A single poster takes one person 30 minutes to create and costs a total of \$1.50.

What would be the total cost for Dorothy to create all the needed posters? Show your work.

If two people working together can create a poster in 20 minutes, how much total time would Dorothy save by getting a friend to help her? Show your work.

If Dorothy works alone for 3 hours and is then joined by her friend, calculate exactly how much total time it will take to create all the necessary posters. Show your work.

Omar, Dorothy's opponent, decided to create his posters on a Saturday and get his friends Janice and Beth to help. He knows that he can create 24 posters in 12 hours if he works alone. He also knows that Janice can create 24 posters in 10 hours and Beth can create 24 posters in 9 hours. How long will it take them, if all three of them work together, to create the 24 posters? Round all decimals to the nearest hundredth. Show your work.

When Omar went to purchase his posters, he discovered that the cost of creating a poster had increased by 20%. How many posters will he be able to create if he wants to spend the same amount of money on his posters as Dorothy? Justify your answer.

LINGUISTICALLY MODIFIED ITEM: You want to plant 6 roses in each of four large pots. Planting a single rose takes you 30 minutes and costs \$1.50.

What is the total cost to plant all the roses? Show your work.

With a friend's help, you can plant a rose in 20 minutes. How much total time do you save by getting a friend to help? Show your work.

You work alone for 3 hours, and then a friend joins you. Now how much total time will it take to plant all the roses? Show your work.

You can plant 24 roses in 12 hours. Your friend Al can plant 24 in 10 hours and your friend Kim can plant 24 in 9 hours. How long does it take the three of you to plant 24 roses together? Round all decimals to the nearest hundredth. Show your work.

You just discovered that the cost of purchasing a rose increased by 20%. How many roses can you plant with the same amount of money that you spent when a rose cost \$1.50? Justify your answer.

Source: Abedi (2010).

Scoring Performance Assessment Tasks

An exemplary performance assessment task, no matter how well written, may not produce desirable outcomes if not scored properly for ELL students of diverse linguistic and cultural backgrounds. Proper scoring of performance assessment tasks is of paramount importance for ELL students, because correct responses in content-based areas (such as math and science) may be masked by potential difficulties in grammar and spelling. Among the components involved in scoring performance assessments are (1) creating a scoring rubric (2) training scorers, and (3) establishing sufficient interscorer reliability. Major attention is needed in all three areas to ensure reliable and valid scoring of a performance assessment.

The first and the most important step to ensure an objective and valid scoring of PATs is access to a well-developed, objective, and validated scoring rubric. A detailed and objective scoring rubric always accompanies a well-designed series of performance tasks because the rubric helps to ensure accurate and consistent scoring. A set of anchor items with the rubric will help operationalize the rubric. Scoring rubrics, particularly in large-scale assessments, are often validated (Jonsson & Svingby, 2007). The process of validating scoring rubrics for use with ELL students should include clear instructions to avoid factors unrelated to the content. For example, in content-based areas such as math and science, where spelling and grammar are not the focal construct of measurement, students should not be penalized for spelling or grammar errors. A scoring rubric developed by the New Jersey SRA illustrates this point (State of New Jersey, 2004). For

Figure 7. A Sample Performance Assessment Task (PAT)

Quadrilateral ABCD is graphed below with A $(-3, 2)$, B $(0, -2)$, C $(8, 4)$, and D $(5, 8)$.

(A) Calculate the slopes of each side of ABCD and of the two diagonals. Show your work and label your responses.

(B) Explain mathematically how you know that the quadrilateral is or is not each of the following types of quadrilateral:

1. parallelogram
2. rectangle
3. rhombus
4. square

Source: New Jersey Department of Education, 2002–03 HSPA/SRA Mathematics, Performance Assessment Task (p. 33, Mathematics).

each PAT, a 4-point scoring rubric is developed with clear performance level descriptors and clear instructions for scoring (see the Appendix).

The next step in objectively scoring PATs is properly training scorers. They should have relevant knowledge and experience before participating in the training sessions. In the training session, the goal of the performance assessment should be clearly discussed and the rubric developed and validated for this assessment should be introduced to the scorers. A set of released PATs could be used for scoring exercise and interscorer reliability can be computed at the scene on site after each round of scoring. Low interscorer reliability should prompt more focused training. Discussion of the issues concerning scoring of PATs for subgroups such as ELL students should be included in the training session.

Establishing interscorer reliability is the final step in objective scoring. Scoring by various individuals with dissimilar backgrounds may be inconsistent if the rubric is not clear enough and more importantly if student background issues that are not directly related to the PATs intervene in the scoring process. For example, for ELLs the students' language background may negatively affect the level of consistency between scorers. More than one scorer should review a sample of the PATs, then the interscorer reliability should be computed using appropriate statistics such as kappa or intraclass correlation (see Abedi, 1996, for discussion of rater reliability coefficients and how to compute them). A low interscorer reliability coefficient (kappa of .50 or below) should call for revision of the rubric and more intense training.

Summary and Discussion

Assessment outcomes have a major impact on the academic career of English language learners because they can influence a student's classification, promotion, and graduation. The outcomes of these assessments are also used for accountability purposes, which may influence academic performance. Unfortunately, research identifies major problems with traditional statewide assessments for these students; linguistic complexity and cultural bias may affect the outcome of assessments, therefore current assessments may not be a valid measure of what these students know and can do (Solano-Flores & Li, 2006; Solano-Flores & Trumbull, 2003; Solano-Flores, 2008).

Performance assessments are a powerful alternative to the traditional standardized achievement test. They can engage ELL students in the assessment tasks and more comprehensively demonstrate their knowledge in content-based areas. These assessments also supply more in-depth information on academic needs and create an environment for students to engage in more cognitively stimulating activities.

Additionally, the outcomes of performance assessments help us understand the nature of the performance gap between ELL and non-ELL students to see whether such a contrast is due to lack of content knowledge or inadequacy in English proficiency. Performance assessments are often thought to be more engaging for students and to better reflect activities inside the classroom. As a result, they increase the level of motivation and effort for all students, particularly for ELLs, who traditionally face challenging situations in taking standardized achievement tests and need more encouragement and support in their academic undertaking. Performance assessments requiring students to evaluate and solve complex problems, conduct research, write extensively, and demonstrate learning in projects are motivating for students in attaining a high level of learning.

Performance assessments can also yield information for diagnostic purposes to assess what students know. They stimulate learning activities and can help teachers decide where to start or determine which groups of students need special attention. Such strategies can monitor processing skills and problem solving as well as competence in particular areas; this is beneficial for ELL students who may not have received equal education opportunities owing to their linguistic needs (Abedi & Herman, 2010). ELL students often exhibit greater interest within the classroom and learn more when required to organize facts around major concepts and actively construct their own understanding of the concepts in a rich variety of contexts.

Performance assessment can contribute to the academic careers of students, particularly those with a challenging academic life, by informing instruction and supporting higher-quality teaching and learning. Such tasks are also instructional, allowing students to actively engage in worthwhile learning activities in the classroom. In performance

assessment settings, students may be encouraged to seek out additional information or try a range of approaches, and in some situations to work in teams. These assessment strategies benefit ELL students because many variables affecting large-scale state and national assessments have less impact on the learning environment of the classroom.

As with many existing standardized achievement tests, performance assessment tasks may suffer from excessive language load. For example, students must read and understand directions to the PATs and contexts in which they are embedded, and write and justify their responses. As illustrated in this paper, performance assessment tasks can be written in a linguistically accessible manner to ensure linguistic complexity is not a source of construct-irrelevant variance, without compromising the content and richness of the assessment items.

References

- Abedi, J. (1996). The Interrater/Test Reliability System (ITRS). *Multivariate Behavioral Research*, 31(4), 409–417.
- Abedi, J. (2006). Language issues in item-development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 377–398). Mahwah, NJ: Erlbaum.
- Abedi, J. (2007). English language proficiency assessment and accountability under NCLB Title III: An overview. In J. Abedi (Ed.), *English language proficiency assessment in the nation: Current status and future practice* (pp. 3–10). Davis: University of California.
- Abedi, J. (2008). Measuring students' level of English proficiency: Educational significance and assessment requirements. *Educational Assessment*, 13(2), 193–214.
- Abedi, J., & Herman, J. L. (2010). Assessing English language learners' opportunity to learn mathematics: Issues and limitations. *Teacher's College Record* 112(3), 723–746.
- Abedi, J., Kao, J., Leon, S., Bayley, R., Ewers, N., Herman, J., et al. (2010). Accessible reading assessments for students with disabilities: The role of cognitive, linguistic, and textual features. Manuscript in preparation.
- Abedi, J., Leon, S., & Kao, J. (2008). *Examining differential distracter functioning in reading assessments for students with disabilities*. (CSE Report No. 743). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Leon, S., & Mirocha, J. (2003). *Impact of students' language background on content-based data: Analyses of extant data* (CSE Report No. 603). Los Angeles: University of California, Center for the Study of Evaluation and National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, 14(3), 219–234.
- Abedi, J., Lord C., & Plummer, J. R. (1997). *Final report of language background as a variable in NAEP mathematics performance* (CSE Report No. 429). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Aguirre-Munoz, Z., Boscardin, C. K., Jones, B., Park, J. E., Chinen, M., Shin, H. S., et al. (2006). *Consequences and validity of performance assessment for English language learners: Integrating academic language and ELL instructional needs into opportunity to learn measures*. (CSE Report No. 678). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Archbald, D. A., & Newmann, F. M. (1988). *Beyond standardized testing: Assessing authentic academic achievement in the secondary school*. Reston, VA: National Association of Secondary School Principals.
- Ayala, C. C., Shavelson, R., & Ayala, M. A. (2001). *On the cognitive interpretation of performance assessment scores*. (CSE Report No. 546). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Bass, K. M., Magone, M. E., & Glaser, R. (2002). *Informing the design of performance assessments using a content-process analysis of two NAEP science tasks*. (CSE Technical Report No. 564). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

- Bauman, J., Boals, T., Cranley, E., Gottlieb, M., & Kenyon D. (2007). Assessing comprehension and communication in English state to state for English language learners (ACCESS for ELLs). In J. Abedi (Ed.), *English language proficiency assessment in the nation: Current status and future practice* (pp.81–91). Davis: University of California.
- Boscardin, C. K., Aguirre-Munoz, Z., Chinen, M., Leon, S., & Shin, H. S. (2004). *Consequences and validity of performance assessment for English learners: Assessing opportunity to learn (OTL) in grade 6 language arts*. (CSE Report No. 635). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Burger, S. E., & Burger, D. L. (1994). Determining the validity of performance-based assessment. *Educational Measurement: Issues and Practice*, 13(1), 9–15.
- Christie, F. (1986). Writing in schools: Generic structures as ways of meaning. In B. Couture (Ed.), *Functional approaches to writing: Research perspectives* (pp.221–239). London: Frances Pinter.
- Christie, F. (2002). The development of abstraction in adolescence in subject English. In M. Schleppegrell & M. C. Colobi (Eds.), *Developing advanced literacy in first and second language: Meaning with power* (pp. 45–66). London: Routledge.
- Chung, G.K.W.K., Delacruz, G. C., & Bewley, W. L. (2006). *Performance assessment models and tools for complex tasks*. (CSE Report No. 682). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, 106(6), 1047–1085.
- Darling-Hammond, L. (2006). No Child Left Behind and high school reform. *Harvard Educational Review*, 76(4), 642–667.
- Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of “No Child Left Behind.” *Race, Ethnicity and Education*, 10(3), 245–260.
- Fiderer, A. (1993). *Teaching writing: A workshop approach*. Scholastic Professional Books. New York: Scholastic, Inc.
- Fiderer, A. (1995). *Practical assessments for literature-based reading classrooms*. Scholastic Professional Books. New York: Scholastic, Inc.
- Fiderer, A. (2009). Performance assessment for reading. Retrieved September 2009 from <http://www.teacher.scholastic.com/professional/assessment/readingassess.htm>.
- Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis*, 20(2), 95–113.
- Goldschmidt, P., Martinez, J. F., Niemi, D., & Baker, E. L. (2007). Relationships among measures as empirical evidence of validity: Incorporating multiple indicators of achievement and school context. *Educational Assessment*, 12(3 & 4), 239–266.
- Heilig, J. V., & Darling-Hammond, L. (2008). Accountability Texas style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis*, 30(2), 75–110.

- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). *Complex, performance-based assessment: Expectations and validation criteria*. (CSE Report No. 331). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Linn, R. L., & Burton, E. (1994). Performance-based assessment: Implications of task specificity. *Educational Measurement: Issues and Practice*, 13(1), 5–8.
- Mehrens, W. A. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice*, 11(1), 3–9.
- Miller, M. D., & Linn, R. L. (2000). Validation of performance-based assessments. *Applied Psychological Measurement*, 24(4), 367–378.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). *Design and analysis in task-based language assessment*. (CSE Report No. 597). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- National Center for Education Statistics (2005). *National assessment of educational progress, 2007, mathematics assessments*. Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved September 2009 from <http://nces.ed.gov/nationsreportcard/itmlsx/search.aspx?subject=mathematics>.
- No Child Left Behind Act*. (2001). Pub. L. No.107–110, §115 Stat. 1425 (2002).
- Parker, C. E., Louie, J., & O’Dwyer, L. (2009). *New measures of English language proficiency and their relationship to performance on large-scale content assessments*. (Issues & Answers Report, REL 2009–No. 066). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands. Retrieved from <http://ies.ed.gov/ncee/edlabs>.
- Shavelson, R., Baxter, G., & Pine, J. (1991). Performance assessment in science. *Applied Measurement in Education*, 4(4), 347–362.
- Shavelson, R., Gao, X., & Baxter, G. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, 30(3), 215–232.
- Shavelson, R., Solano-Flores, G., & Ruiz-Primo, M. (1998). Toward a science performance assessment technology. *Evaluation and Program Planning*, 21(2), 171–184.
- Shepard, L. A. (1991). Psychometricians’ beliefs about learning. *Educational Researcher*, 20(7), 2–16.
- Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English language learners. *Educational Researcher*, 37(4), 189–199.
- Solano-Flores, G., & Li, M. (2006). The use of Generalizability (G) Theory in the testing of linguistic minorities. *Educational Measurement: Issues and Practice*, 25(1), 13–22.
- Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher*, 32(2), 3–13.

State of New Jersey (2004). *Mathematics: A rubric scoring handbook*. (PTM# 1504.30) [Special Review Assessment, High School Proficiency Assessment]. Trenton: Department of Education.

Wang, J., Niemi, D., & Wang, H. (2007a). *Impact of different performance assessment cut scores on student promotion*. (CSE Report No. 719). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Wang, J., Niemi, D., & Wang, H. (2007b). *Predictive validity of an English language arts performance assessment*. (CSE Report No. 729) Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Appendix: Sample PAT Response and Rubric

CLUSTER 1/MACRO B
RUBRIC-B0705-001R

Page 1 of 2

2002-2003 PRACTICE MATERIALS
New Jersey Department of Education
HSPA/SRA MATHEMATICS
Sample Response & Scoring Rubric

Sample Response:

- A) $(\$1.50) \times (4 \times 6) = \mathbf{\$36}$
- B) $(30 \times 24) - (20 \times 24) = 240 \text{ minutes} = \mathbf{4 \text{ hours}}$
or $24 \times 10 \text{ (time saved per poster)} = \mathbf{240 \text{ minutes}}$
 $(3)(4) - 2(4) = 12 - 8 = \mathbf{4 \text{ hours}}$
- C) One person can make 6 posters in 3 hours. The remaining 18 posters take 2 people 360 minutes or 6 hours. $3 \text{ hours} + 6 \text{ hours} = \mathbf{9 \text{ hours total}}$.
- D) $\frac{h}{12} + \frac{h}{10} + \frac{h}{9} = 1$
 $.29\bar{h} = 1$
 $h = \mathbf{3.45 \text{ hours}}$ (note: used rounding process twice)
 $h = \mathbf{3.40 \text{ hours}}$ (note: used rounding process only once)
OR
 $180 \left(\frac{h}{12} + \frac{h}{10} + \frac{h}{9} \right) = (180)$
 $15h + 18h + 20h = 180$
 $53h = 180$
 $h = \frac{180}{53} \text{ hours} = 3 \frac{21}{53} \text{ hours}$
or $\mathbf{3.40 \text{ hours}}$
- E) $(1.20)(1.50) = \$1.80 \text{ per poster}$
 $\$36 \div 1.80 = \mathbf{20 \text{ posters}}$
Omar can create 20 posters.

- 3 Points** The student found the correct total cost for creating the posters, the amount of time saved if a friend helps, and the correct total time needed to create all posters in part C. The total time for Omar to create 24 posters is correctly calculated in part D, and “20” is the correct answer for part E. Work was shown for all parts.
- 2 Points** The student found three or four correct answers out of five, and supported those answers with appropriate work. The other answers were either omitted, incorrect due to a process error, or not supported by work.
- 1 Point** The student found one or two correct answers out of five, supported with appropriate work.
OR
The student found at least three correct answers, but no work was shown.
- 0 Points** The response shows an insufficient understanding of the problem’s essential mathematical concepts. The procedures, if any, contain major errors. There may be no explanation of the solution or the reader may not be able to understand the explanation. The reader may not be able to understand how and why decisions were made.

1. $6 \times 4 = 24$

$24 \times 1.50 = \$36 //$

2. 2 persons 1 passer in 20 minutes

24

$480 = 8 \text{ hours}$

12-

1 person 1 passer 30 minutes

8

24

$720 = 12 \text{ hours}$

4 hours //

3. 1 person 2 passers 180 minutes

$180 : 30 = 6 //$

6

$720 \text{ min. } 12 \text{ hours}$

$180 \text{ min. } 3 \text{ hours}$

$540 \text{ min. } 9 \text{ hours}$

Janice 24 passers in 10 hours 2.4

Beth 24 passers in 9 hours 2.7

Anna 24 passers in 12 hours 2

3 persons in

Linda Darling-Hammond, Co-Director
Stanford University Charles E. Ducommun Professor of Education

Prudence Carter, Co-Director
*Stanford University Associate Professor of Education and
(by courtesy) Sociology*

Carol Campbell, Executive Director



Stanford Center for Opportunity Policy in Education
Barnum Center, 505 Lasuen Mall
Stanford, California 94305
Phone: 650.725.8600
scope@stanford.edu

<http://edpolicy.stanford.edu>