UCLA | CRESST NATIONAL CENTER FOR RESEARCH ON EVALUATION, STANDARDS, AND STUDENT TESTING

USING FEATURE ANALYSIS TO EXAMINE CAREER READINESS IN HIGH SCHOOL ASSESSMENTS

Jenny C. Kao, Kilchan Choi, Nichole M. Rivera, Ayesha Madni, and Li Cai

DECEMBER 2018

CRESST REPORT **858**

The work reported herein was supported under an Enhanced Assessment Grant, PR/Award #S368A150011, with funding to the California Department of Education (CDE), and subaward to the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

The findings and opinions expressed here do not necessarily reflect the positions or policies of the California Department of Education or the U.S. Department of Education.

To cite from this report, please use the following as your APA reference: Kao, J. C., Choi, K., Rivera, N. M., Madni, A. & Cai, L. (2018). *Using feature analysis to examine career readiness in high school assessments* (CRESST Report 858). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Table of Contents

Using Feature Analysis to Examine Career Readiness in High School Assessments[1]

Jenny C. Kao, Kilchan Choi, Nichole M. Rivera, Ayesha Madni, and Li Cai

CRESST/University of California, Los Angeles

Abstract: This report is the second in a series considering career-readiness features within high school assessments. Experts in English language arts and math were trained to rate a selection of active Grade 8 and Grade 11 Smarter Balanced items using feature set lists that were refined within the items' respective content areas (30 features for ELA items; 22 features for math items). A total of 264 ELA items and 186 math items were rated. ELA items contained between three and 13 career-readiness features, with an average of 6.0. The most frequent features were importance of being exact or accurate, written comprehension, time sharing, deductive reasoning, and reading comprehension. Math items contained between two and 15 career-readiness features, with an average of 7.8. The most frequent features were deductive reasoning, analyzing data or information, reading comprehension, number facility, and processing information. Feature ratings of the target items were analyzed with item metadata difficulty parameters in order to explore relationships between features and item difficulty. A number of career-readiness features showed associations with item difficulty, notably, reading comprehension for Grade 8 math and deductive reasoning for Grade 11 ELA. Because career-readiness features can be used to explain item difficulty, results suggest that such features are prevalent in content-based assessments, and inferences for career readiness can thus be drawn from test performance.

Introduction

By 2020, it is expected that 65% of all jobs will require some postsecondary education, with only 36% of jobs open to workers with only a high school education (Carnevale, Smith, & Strohl, 2013). In 2017, about 16.3 million people ages 16 to 24 (or, 42.7% of the total) were not enrolled in any school. However, national high school graduation rates are at an all-time high (EDFacts, 2018), and about two thirds of high school graduates enroll in colleges or universities

[1]We thank Eric Zilbert and Kelly Bacher of the California Department of Education for their support, guidance, and helpful comments. We are grateful to Eva Baker and Noelle Griffin for their advice and support. We also thank Sandy Chang, Markus Iseli, Eunhee Keum, Junok Kim, Alan Koenig, and Julia Phelan for their contributions to this research. Special thanks to Robert Mislevy and David Thissen for reviewing early drafts of this report.

(Bureau of Labor Statistics, 2018). Considering college and career readiness for high school students is critical, especially at a time when classrooms are still largely unconnected to the workplace and success in postsecondary education. Higher levels of education unfortunately do not necessarily guarantee college or career readiness. Roughly eight in 10 students entering community college in California require at least one developmental course in math or English (Mejia, Rodriguez, & Johnson, 2016). The state of California has been part of a consortium of states implementing Smarter Balanced, a K-12 assessment system designed to be aligned with college- and career-ready standards, and built with higher education in mind. Smarter Balanced test scores in English language arts (ELA) and math are accepted by over 200 colleges and universities in 10 states to determine whether students can be exempted from remedial courses, while the state of South Dakota uses Smarter Balanced scores for admission to its public universities (Smarter Balanced, 2017).

In 2017, the Bureau of Labor Statistics projected the fastest growing occupations between 2016 and 2026 will grow about 7%, with larger growth in highly specialized areas like solar photovoltaic installers, wind turbine service technicians, and nurse practitioners (Bureau of Labor Statistics, 2017). While these jobs may require highly specialized skills, they may also require some more general skills, not often directly addressed in standardized testing. For example, 96% of all jobs consider critical thinking and active listening as critical to success (Carnevale et al., 2013). Such general skills may be addressed in ELA and/or math test items. If so, then we may be able to draw inferences about students' career readiness based on their test scores.

This report is the second in a series of reports considering career-readiness features within high school assessments. In our first report (Madni, Kao, Rivera, Baker, & Cai, 2018), we discussed the identification and refinement of a set of career-readiness features for standardized assessments. A test is valid in the context of the thing it is validating. If assessments reflect necessary criteria for career and college success, including cognitive and noncognitive skill demands, we may see an additional pathway for K–12 assessments to offer long-term value. Assessments that accurately project college and career readiness may potentially impact the readiness of the workforce as a whole. This report focuses on a feature analysis aimed toward developing and refining the career-readiness feature set for standardized assessments. Some literature and discussion of both feature analysis and the development of the career-readiness feature set from our first report are repeated here for context.

Feature Analysis: Theoretical Background

Baker, Madni, Michiuye, Choi, and Cai (2015) noted that one of the earliest references to the idea of feature analysis can be found in the *Report of the Commission on Tests: II* (Gordon, 1970). Gordon mentioned qualitative analysis of assessments should emphasize "description and prescription," that is, the qualitative *description* of cognitive functions leading to the

prescription of the learning experiences required to more adequately ensure academic success. Gordon recommended that the College Examination Board add "descriptive patterns of achievement and function" to its assessment reports. Gordon suggested that existing instruments can be examined with a view towards categorization and interpretation to determine whether data can be reported in qualitative ways, to supplement traditional quantitative reports. As Gordon and Rajagopalan note almost 50 years later, "assessment *of* education" is not the singular goal – instead, "assessment [is] *for* education… in the service of teaching and learning" (Gordon & Rajagopalan, 2016, p.11).

The main rating framework underlying the feature analysis work described here is derived from Baker and O'Neil's (2002) approach to designing problem-solving assessments, Jonassen's (2000) typology of problems, and CRESST's problem solving ontology. Jonassen's work on problem solving articulated problem types vary on "at least" three different dimensions: problem type, problem representation, and individual differences (p. 72). Problem types vary in "structuredness," abstractness (domain specificity), and complexity. Problem representation refers to the context, cues or clues, and modality surrounding a problem and individual differences in the problem solver.

Baker and O'Neil's approach characterizes three types of problem-solving tasks: (a) a task in which an appropriate solution is known in advance, (b) a task in which there is no known solution to the problem, and (c) a task that requires the application of a given tool set to a broad range of topics. Unlike previous models, this approach provides criterion-referenced evidence in support of assessment validity claims by integrating feature rating and step-by-step analysis with modern statistical techniques. These are all features of problem solving that can be rated as part of a particular item and that are associated with a particular cognitive demand or process on the part of the student. Identifying the problem is often one of the most difficult aspects of problem solving (see Baker & O'Neil, 2002). The ambiguity of problem identification may be dependent on the prior knowledge of the learner and the purpose of the assessment. An assessment developer can adjust the difficulty of a task or item by stating the problem explicitly or obscuring it in context, such as within a narrative. The difficulty of an item or task can also be adjusted by either providing extraneous information or developing a task with missing information that needs to be constructed. These types of adjustments might increase difficulty and also require an associated cognitive demand or cognitive process that might be more complex.

CRESST has previously used qualitative content analysis techniques to assess language demands in standardized assessments for math, ELA, and tests of English language proficiency (Wolf et al., 2008; Wolf, Kim, Kao, & Rivera, 2009) as well as cognitive, grammatical, textual, and visual features on standardized assessments for students with disabilities (Abedi et al., 2011, 2012). A quantitative technique similar to feature analysis was previously introduced in Roberts, Chung, and Parks (2016) and used to categorize attributes of metadata created when children interacted with educational online games and media. This approach was previously

used in work with PBS (Chung & Parks, 2015; Chung & Redman, 2015a, 2015b). This model ensures validity by going beyond simple task descriptions, and by yielding explanation for possible areas of growth, identifying task elements that are suitable for instruction, and lastly, providing a method for comparability and prediction.

Goals of the Present Study

This report focuses on a feature analysis aimed toward developing a career-readiness feature set for standardized assessments. Feature analysis is defined as the qualitative rating of tasks against a set of attributes, followed by a subsequent quantitative analysis to determine how these attributes contribute to task performance. The overall feature analysis process includes feature rating, step-by-step analysis, cognitive labs, and quantitative analysis (Baker et al., 2015). Feature analysis, in the context of career readiness, aims to address such questions as (a) What particular attributes/features does each item contain? (b) Which features appear more frequently across items? (c) Are there differences in feature representation across domain areas (i.e., math vs. ELA)? and (d) Is there a relationship between the features within an item and the item's difficulty?

Our prior report, which used cognitive laboratory interviews, explored some of these questions (Madni et al., 2018). Cognitive lab interviews were conducted with 17 high school students using six Smarter Balanced practice test items (four math and two ELA) following a preliminary feature rating of the items. Results indicated that each test item contained between eight and 13 career-readiness features. The cognitive labs provided preliminary evidence that career-readiness features can indeed be found in a small and targeted set of practice test items, which paved the way for the larger study. The present report focuses on the larger study, in which content-area experts were recruited to refine the list of features and rate a larger number of active test items. Results of feature ratings and quantitative analyses with recent student performance data are presented in this report.

Feature Set List Creation

As discussed in Madni et al. (2018), the goal of the feature rating or scoring process is to determine what features and attributes are present or absent in a particular item, and what steps need to be completed to solve an item or task correctly, and to perform descriptive analyses across features and items. This task is performed by content-area experts. The features are refined to target a particular content area prior to feature rating by the experts.

The feature creation process requires several steps. The initial step involves selecting and reviewing key resource materials. To create the current set of career skills and features, researchers reviewed the Bureau of Labor Statistics career data, the O*NET online databases, and previous CRESST ontologies and feature and rating schemes. These resource materials were studied to determine an initial set of career skills. These skills were refined by selecting those that were categorized as most important based on O*NET importance ratings. These two

careers are emergency medical technician (EMT), representing the healthcare industry, and web developer, representing the technology industry. These careers were chosen as exemplars to represent two different industries, and because both careers, at the time of the study, were reported by the Bureau of Labor Statistics as having "Bright Outlook" and "rapid growth." In addition, these two careers did not require a four-year degree, but did require some postsecondary training.

This feature set was further reduced by utilizing Smarter Balanced items and blueprints as selection criteria (i.e., features that were not likely to be found within the items were taken out). Previous CRESST feature analysis results also informed the current feature set. Specifically, features that were found to contribute variance in previous CRESST studies were included as part of the current set. Finally, the feature set was refined by incorporating the expertise of select subject-matter experts (SMEs) in college and career readiness, business, pre-hospital care, and web development. These SMEs filled out a survey where they answered targeted questions about the career feature set.

The SMEs were first asked to indicate the extent to which each career skill was not applicable, contributes to, or was essential to effectively and successfully complete daily job-related tasks. The SMEs then indicated which 30 skills were most important from those rated as "essential." The SMEs were then asked to rank these 30 skills in order of importance and create an operational example of the 15 highest ranked skills. After this feature selection process, the SMEs were asked to review and verify the final set that would be utilized for feature rating. The goal was for the features to be action-oriented with adequate granularity to allow for implementation with low inference across domains and task types.

Target Career-Readiness Features

Table 1 shows the initial set of 36 career-readiness features (across both exemplar careers) created and reviewed by SMEs following the process delineated above. They are grouped broadly into three categories: Skills, Abilities, and Work Activities/Context for ease of presentation. "Skills" generally refer to developed capacities that facilitate learning or the more rapid acquisition of knowledge. "Abilities" generally refer to enduring attributes of the individual that influence performance. "Work Activities" generally refer to the types of job behaviors occurring on multiple jobs. "Work Context" refers to the physical and social factors that influence the nature of the work. The Method section describes how the feature set was further refined for rating both ELA and math test items.

Table 1

Target Career-Readiness Features by Category

Feature	Description
Features related to skills	
Active learning	Understanding the implications of new information for both current and future problem solving and decision making.
Active listening	Giving full attention to what other people are saying and taking time to understand the points being made.
Complex problem solving	Identifying complex problems and reviewing related information to develop and evaluate options and implement solutions.
Critical thinking	Using logic and reasoning to identify the strengths and weaknesses of alternative solutions, conclusions, or approaches to problems.
Judgment and decision making	Considering the relative costs and benefits of potential actions to choose the most appropriate one.
Mathematics	Using mathematics to solve problems.
Monitoring	Monitoring/assessing performance of yourself, other individuals, or organizations to make improvements or take corrective action.
Reading comprehension	Understanding written sentences and paragraphs in work-related documents.
Features related to abilities	
Deductive reasoning	The ability to apply general rules to specific problems to produce answers that make sense.
Flexibility of closure	The ability to identify or detect a known pattern, figure, object, word, or sound that is hidden in other distracting material.
Fluency of ideas	The ability to come up with a number of ideas about a topic.
Inductive reasoning	The ability to combine pieces of information to form general rules or conclusions.
Information ordering	The ability to arrange things or actions in a certain order or pattern according to a specific rule or set of rules.
Mathematical reasoning	The ability to choose the right mathematical methods or formulas to solve a problem.
Memorization	The ability to remember information such as words, numbers, pictures, and procedures.
Number facility	The ability to add, subtract, multiply, or divide quickly and correctly.
Oral comprehension	The ability to listen to and understand information and ideas presented through spoken words and sentences.

Feature	Description
Problem sensitivity	The ability to tell when something is wrong or is likely wrong. It does not involve solving the problem, only recognizing there is a problem.
Selective attention	The ability to concentrate on a task over a period of time without being distracted.
Time sharing	The ability to shift back and forth between two or more activities or sources of information.
Written comprehension	The ability to read and understand information and ideas presented in writing.
Written expression	The ability to communicate information and ideas in writing so others will understand.
Visualization	The ability to imagine how something will look after it is moved around or when its parts are moved or rearranged.
Features related to work activities/context	
Analyzing data and information	Identifying the underlying principles, reasons, or facts of information by breaking down information or data into separate parts.
Documenting/recording information	Entering, transcribing, recording, storing, or maintaining information in written or electronic/magnetic form.
Estimating the quantifiable characteristics of products, events, or information	Estimating sizes, distances, and quantities; or determining time, costs, resources, or materials needed to perform a work activity.
Getting information	Observing, receiving, and otherwise obtaining information from all relevant sources.
Identifying objects, actions, and events	Identifying information by categorizing, estimating, recognizing differences or similarities, and detecting changes in circumstances or events.
Importance of being exact or accurate	Being very exact or highly accurate is important to performing this job.
Interacting with computers	Using computers and computer systems, including hardware and software, to program, write software, set up functions, enter data, or process information.
Judging the qualities of things, services, or people	Assessing the value, importance, or quality of things or people.
Making decisions and solving problems	Analyzing information and evaluating results to choose the best solution and solve problems.
Organizing, planning, and prioritizing work	Developing specific goals and plans to prioritize, organize, and accomplish your work.
Processing information	Compiling, coding, categorizing, calculating, tabulating, auditing, or verifying information or data.

Feature	Description
Thinking creatively	Developing, designing, or creating new applications, ideas, relationships, systems, or products, including artistic contributions.
Updating and using relevant knowledge	Keeping up-to-date technically and applying new knowledge.

Method

Feature Rating

Preliminary steps. CRESST researchers met to discuss and reduce the number of features in the original set. The goal was to optimize the list for rating ELA and math items while also reducing measurement error and increasing analytical power. To reduce the feature list researchers met and independently rated the feature list for ELA only. We considered each feature in regard to whether an ELA item would exhibit or consist of that feature. For instance, mathematics, mathematical reasoning, and number facility were not relevant to ELA and omitted. The process was repeated for math. Researchers then reconvened to discuss disagreements until a consensus was reached. Two separate feature sets—one for ELA and one for math—were subsequently produced.

From the original set of 36, six features were removed from the ELA set and five features were removed from the math set. The six features removed from the feature list for ELA were mathematics; mathematical reasoning; number facility; selective attention; estimating the quantifiable characteristics of products, events, or information; and updating and using relevant knowledge. Five features were removed from the feature list for math: active listening, judgment and decision making, oral comprehension, selective attention, and updating and using relevant knowledge. CRESST researchers deemed such features as not relevant to the respective content-area test items.

Content area experts. Three experts were recruited for each content area. The ELA experts held doctorates in education and literacy. The math experts held doctorates or master's degrees related to engineering, measurement, and statistics. The experts met and trained separately and the math experts further reduced the feature set list as deemed relevant to math assessment items. Math raters also removed features deemed redundant for rating math items (i.e., mathematical reasoning was retained in favor of mathematics; reading comprehension was retained in favor of written comprehension). This led to two separate sets of features: a set of 30 features for ELA items, and a set of 22 features for math items. Table 2 shows the feature set by content area, with "Y" indicating that the feature was included in the final set prior to rating for each content area.

Table 2

Set of Career-Readiness Features Used for Rating Test Items by Content Area

Related area	Feature	ELA	Math
Skills	Active learning	Y	Y
Skills	Active listening	Y	
Skills	Complex problem solving	Y	Y
Skills	Critical thinking	Y	Y
Skills	Judgment and decision making	Y	
Skills	Mathematics		
Skills	Monitoring	Y	
Skills	Reading comprehension	Y	Y
Abilities	Deductive reasoning	Y	Y
Abilities	Flexibility of closure	Y	Y
Abilities	Fluency of ideas	Y	
Abilities	Inductive reasoning	Y	Y
Abilities	Information ordering	Y	Y
Abilities	Mathematical reasoning		Y
Abilities	Memorization	Y	Y
Abilities	Number facility		Y
Abilities	Oral comprehension	Y	
Abilities	Problem sensitivity	Y	
Abilities	Selective attention		
Abilities	Time sharing	Y	Y
Abilities	Written comprehension	Y	
Abilities	Written expression	Y	Y
Abilities	Visualization	Y	Y
Work activities/context	Analyzing data and information	Y	Y
Work activities/context	Documenting/recording information	Y	
Work activities/context	Estimating the quantifiable characteristics of products, events, or information		Y
Work activities/context	Getting information	Y	Y
Work activities/context	Identifying objects, actions, and events	Y	Y
Work activities/context	Importance of being exact or accurate	Y	Y

Related area	Feature	ELA	Math
Work activities/context	Interacting with computers	Y	
Work activities/context	Judging the qualities of things, services, or people	Y	
Work activities/context	Making decisions and solving problems	Y	Y
Work activities/context	Organizing, planning, and prioritizing work		Y
Work activities/context	Processing information	Y	Y
Work activities/context	Thinking creatively	Y	
Work activities/context	Updating and using relevant knowledge		
	Total	30	22

Rater training. Content-area experts were trained by CRESST researchers on the feature set for respective content areas by rating and discussing a variety of test items. Operational definitions were refined by the experts for each feature. Each feature was rated on a scale of 1 to 4, with a rating of 1 generally referring to little or no presence of the feature (in order to solve the problem), and 4 generally referring to the feature being present or necessary to solve the problem. ELA experts continued to meet together until achieving an acceptable interrater reliability agreement (above .80). The math experts also continued to meet together until achieving acceptable interrater reliability. The experts then rated a selection of active test items (described below) on their own. About 20% of the items were randomly selected to be rated by more than one rater to compute interrater reliability. The interrater reliability (based on percent agreement) was .88 for ELA items and .85 for math items.

Test Items

Initially, items to be rated were proportionally selected from the California pool of Smarter Balanced items based on item types, claims, and targets across the two content areas across two school years (2015 and 2016) and across two grade levels (Grade 8 and Grade 11). (For more information regarding Smarter Balanced item types, claims, and targets, please see www.smarterbalanced.org.) Only items from the summative computer-adaptive test were selected. Eighty items were selected for ELA from each year and from each grade level for a total of 320 items. However, because fewer items were available for Grade 8 math, 80 items were selected for math from each year for Grade 11, while only 60 items were selected from each year for Grade 8, for a total of 280 items. Because some of the items appeared across both years, there were a total of 264 unique ELA items and 186 unique math items.

Analysis Plan

Descriptive statistics of the feature ratings were computed for both ELA and math items across both grade levels. Feature ratings were recoded into binary codes, with ratings of 1 and

2 recoded as 0 (feature is not present) and ratings of 3 and 4 recoded as 1 (feature is present). A random sampling of student performance data from the state of California were used for preliminary analyses. Five thousand cases were randomly sampled from both Grade 8 and from Grade 11, and from two different test administration years.

Results

English Language Arts

Descriptive results. Thirty career-readiness features were used for rating Smarter Balanced items. Of the 264 ELA items, 126 items were from Grade 8 and 136 items were from Grade 11 (two items were coded as both grade levels and were not included in grade-level analyses). Of the 264 ELA items, 152 of the items were from the 2015 test administration and 158 items were from the 2016 test administration. Feature ratings were recoded into binary codes, so that all ratings of 1 and 2 were recoded as 0, and all ratings of 3 and 4 were recoded as 1. Based on the recoded data, ELA items contained between three and 13 career-readiness features, with an average of 5.97. Table 3 shows the means and standard deviations of the original 4-point ratings. Table 3 also shows the frequency and percentage of items recoded as 1 (present) for each career-readiness feature.

Table 3

Career-Readiness Features for ELA Items (264 Items)

Feature	Original ratings (scale of 1 to 4)		Recoded ratings (0 or 1)	
	Mean	*SD*	Frequency	Percentage
Features related to skills				
Reading comprehension	3.2	1.3	190	72.0
Active listening	1.6	1.2	57	21.6
Critical thinking	1.9	0.3	1	0.4
Complex problem solving	1.3	0.5	0	0.0
Active learning	1.0	0.0	0	0.0
Judgment and decision making	1.0	0.0	0	0.0
Monitoring	1.0	0.0	0	0.0
Features related to abilities				
Written comprehension	4.0	0.0	264	100.0
Time sharing	2.9	0.6	223	84.5
Deductive reasoning	2.8	0.5	209	79.2
Inductive reasoning	2.3	0.8	114	43.2
Oral comprehension	1.6	1.2	57	21.6
Flexibility of closure	1.8	0.7	32	12.1
Information ordering	1.2	0.6	23	8.7
Written expression	1.3	0.8	22	8.3
Fluency of ideas	1.2	0.6	22	8.3
Memorization	1.9	0.3	0	0.0
Problem sensitivity	1.1	0.2	0	0.0
Visualization	1.0	0.0	0	0.0
Features related to work activities/context				
Importance of being exact or accurate	3.9	0.3	264	100.0
Identifying objects, actions, and events	2.1	0.3	26	9.8
Documenting/recording information	1.3	0.8	23	8.7
Analyzing data or information	1.5	0.8	20	7.6
Getting information	2.0	0.3	14	5.3
Making decisions and solving problems	2.0	0.3	6	2.3

Feature	Original ratings (scale of 1 to 4)		Recoded ratings (0 or 1)	
	Mean	*SD*	Frequency	Percentage
Judging the qualities of things, services, or people	1.9	0.4	4	1.5
Thinking creatively	1.1	0.3	3	1.1
Interacting with computers	1.3	0.5	1	0.4
Organizing, planning, and prioritizing work	1.0	0.0	0	0.0
Processing information	1.9	0.4	0	0.0

Note. Rows are sorted by percentages, in descending order, within each category.

Among the seven features related to skills, the more common features in ELA items were reading comprehension and active listening. Among the 12 features related to abilities, the more common features were written comprehension, time sharing, and deductive reasoning. Among the 11 features related to work activities or context, importance of being exact or accurate was prominent. In the 4-point ratings, five features were rated as 1 (not present) for all items: active learning; judgment and decision making; monitoring; visualization; and organizing, planning, and prioritizing work. An additional four features were also frequently rated as 1, with no ratings higher than 2: complex problem solving, memorization, problem sensitivity, and processing information. Patterns of the number of features rated as present in the items were similar across the two grade levels.

The remaining results for ELA items involve analyses using the 2-point scale. Table 4 below shows the total number of features rated as present (i.e., rated as 1) in ELA items by feature category.

Table 4

Total Number of Features Rated as Present in ELA Items by Feature Category

Number of features	Features related to skills		Features related to abilities		Features related to work activities/context	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
0	19	7.2	–	–	–	–
1	242	91.7	17	6.4	217	82.2
2	3	1.1	42	15.9	25	9.5
3	–	–	78	29.6	–	–
4	–	–	62	23.5	16	6.1
5	–	–	30	11.4	6	2.3
6	–	–	16	6.1	–	–
7	–	–	15	5.7	–	–
8	–	–	4	1.5	–	–
Total	264	100.0	264	100.0	264	100.0

Table 5 shows the correlation matrix of the rating of each feature for ELA items. The matrix shows the extent to which a feature was correlated with other features. As expected, the presence of the active listening feature is negatively correlated with the presence of the reading comprehension feature. Fluency of ideas correlated highly with information ordering (corr = .98) as well as documenting/recording information (corr = .98). Information ordering correlated highly with written expression (corr = .98). Some features had positive but low range of correlations (approximately .3 in magnitude). For example, the presence of inductive reasoning is positively correlated with the presence of deductive reasoning; flexibility of closure; fluency of ideas; information ordering; oral comprehension; timesharing; analyzing data or information; documenting/recording information; and identifying objects, actions, and events. The presence of the deductive reasoning feature also shows a positive correlation with these features, but the magnitude of the correlation coefficients are much smaller (approximately .1–.2). Note that the correlations of some features which do not have variability were not included. This includes features that were coded as 1 (i.e., present in all of the items): the importance of being exact or accurate. It also includes features that were coded as 0 (i.e., not present in any of the items): active learning; complex problem solving; judgment and decision making; monitoring; memorization; problem sensitivity; visualization; organizing, planning, and prioritizing work; and processing information.

Table 5

Correlation Matrix of Features Rated as Present in ELA Items

Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Active listening	–																	
2. Critical thinking	-.03	–																
3. Reading comprehension	**-.80**	.04	–															
4. Deductive reasoning	.09	.03	**.18**	–														
5. Flexibility of closure	**.23**	**.17**	**-.16**	**.16**	–													
6. Fluency of ideas	**-.16**	-.02	**.19**	**.12**	.10	–												
7. Inductive reasoning	**.23**	.07	-.05	**.33**	**.29**	**.29**	–											
8. Information ordering	**-.16**	-.02	**.19**	**.13**	.09	**.98**	**.30**	–										
9. Oral comprehension	**1.00**	-.03	**-.80**	.09	**.23**	**-.16**	**.23**	**-.16**	–									
10. Time sharing	.10	.03	**.17**	**.22**	.10	**.13**	**.25**	**.13**	.10	–								
11. Written expression	**-.16**	-.02	**.19**	**.12**	.10	**.95**	**.29**	**.98**	**-.16**	**.13**	–							
12. Analyzing data or information	**-.15**	-.02	**.18**	.11	.11	**.90**	**.30**	**.93**	**-.15**	**.12**	**.90**	–						
13. Documenting/recording information	**-.16**	-.02	**.19**	**.13**	.09	**.98**	**.30**	1.00	**-.16**	**.13**	**.98**	**.93**	–					
14. Getting information	.04	-.01	.03	.08	-.04	.23	.07	.23	.04	.10	.23	.12	.23	–				
15. Identifying objects, actions, and events	-.08	-.02	**.12**	**.11**	**.15**	**.87**	**.33**	**.89**	-.08	**.14**	**.87**	**.87**	**.89**	**.21**	–			
16. Interacting with computers	-.03	.00	.04	.03	-.02	-.02	-.05	-.02	-.03	.03	-.02	-.02	-.02	-.01	-.02	–		
17. Judging qualities of things, services, or people	-.07	-.01	.08	.06	-.05	-.04	-.11	-.04	-.07	.05	-.04	-.04	-.04	-.03	-.04	-.01	–	
18. Making decisions and solving problems	-.08	.40	.10	.08	.10	-.05	-.08	-.05	-.08	.07	-.05	-.04	-.05	-.04	-.05	-.01	-.02	–
19. Thinking creatively	-.06	-.01	.07	-.03	.07	.36	.05	.35	-.06	.05	.36	.37	.35	-.03	.32	-.01	-.01	-.02

Note. Features with no variability (i.e., either not present in any item or present in all items) were excluded from correlations. Bold indicates significance at $p < .05$.

Relationship between career-readiness features and item difficulty for ELA. One of the goals of the present study was to explore whether the career-readiness features of a test item were associated with the item's difficulty. In other words, the presence of certain career-readiness feature(s) might make an item easier or harder. There are two important aspects to consider when addressing this question. First, as Smarter Balanced was administered through computer-adaptive testing (CAT), each student was tested with different items depending upon calibrated ability. Thus, a random sampling of student performance data may not yield all of the rated items. For example, we randomly sampled 5,000 students from the 2015 Grade 8 test administration. We examined how many of the 80 items rated for 2015 Grade 8 were administered to the 5,000 sampled. Some students had only one item among the 80 items, while some students had as many as 16 items. On average, about seven to eight items per student were administered in this random sampling.

Second, given the sparse information on item response per student by design (i.e., the small number of items selected—80 items among the large item pool of more than 1,000 items) and the CAT aspect of item delivery, care needs to be taken when using a linear logistic test model (LLTM) or a crossed random item model. The item parameters, for example item difficulty, estimated using the sampled items and students are different from the metadata item parameters that were estimated using the whole test population data. Thus, employing either an LLTM or a crossed random item model using the sampled data would lead to inappropriate results unless the metadata item parameters were superimposed with the sample-based item parameters.

As an alternative, we employed a multiple regression method using the metadata item difficulty parameter as the outcome and a set of career-readiness features as covariates using a weighted least squares (WLS) estimator, more specifically, var (Y_i) = σ_i^2, where $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ are known error variance from calibration and σ_i^2 = variance of item difficulty parameter for item *i*. Weight (w) = $1/\sigma_i^2$, thus, items with larger error variance put less weight in a regression fitting. The variance of item difficulty equals the square of standard errors of item difficulty / *N* of items in the item pool.

Table 6 shows the estimates, estimation errors, and *p* values from the analyses of Grade 8 ELA items. The expected item difficulty (i.e., intercept) when no features in the model are present is about 1.0. The estimate of each variable shows the expected increase or decrease of the item difficulty parameter when the particular career-readiness feature is present.

Table 6

Relationships Between Item Difficulty Estimate and Career-Readiness Features: Results From Weighted Least Squares Regression for Grade 8 ELA

Variable	Estimate	*SE*	*t*	*p*
Intercept	1.028	0.259	3.97	.000
Fluency of ideas*	1.233	0.335	3.68	.000
Deductive reasoning	0.215	0.220	0.98	.330
Flexibility of closure	0.116	0.162	0.71	.477
Analyzing data or information	0.010	0.691	0.01	.989
Identifying objects, actions, and events	-0.002	0.692	0.00	.998
Getting information	-0.030	0.498	-0.06	.952
Time sharing	-0.051	0.145	-0.35	.727
Inductive reasoning	-0.070	0.187	-0.38	.708
Reading comprehension*	-0.568	0.198	-2.87	.005
Active listening*	-1.180	0.249	-4.74	<.0001

Note. R square = .56. Features are sorted by coefficients, from positive to negative.
*$p < .05$.

The statistically significant coefficient of fluency of ideas was positive, which suggests that the presence of fluency of ideas, while holding other features constant, leads to an increase in item difficulty. The statistically significant negative coefficients of active listening and reading comprehension suggest that the presence of these features leads to a decrease in item difficulty.

Table 7 presents similar results for Grade 11 ELA items. The expected item difficulty (i.e., intercept) when all the features in the model are not present is about 0.6.

Table 7

Relationships Between Item Difficulty Estimate and Career-Readiness Features: Results From Weighted Least Squares Regression for Grade 11 ELA

Variable	Estimate	*SE*	*t*	*p*
Intercept	0.558	0.296	1.88	.062
Fluency of ideas	1.266	0.680	1.86	.065
Identifying objects, actions, and events*	1.219	0.435	2.8	.006
Making decisions and solving problems	0.848	0.777	1.09	.277
Deductive reasoning*	0.659	0.141	4.68	<.0001
Time sharing	0.079	0.170	0.47	.641
Written expression	0.075	0.434	0.17	.863
Getting information	0.056	0.364	0.15	.878
Inductive reasoning	-0.048	0.139	-0.34	.732
Active listening	-0.321	0.344	-0.93	.353
Reading comprehension	-0.331	0.328	-1.01	.314
Flexibility of closure*	-0.961	0.125	-7.67	<.0001
Information ordering	-1.265	0.911	-1.39	.167

Note. R square = .54. Features are sorted by coefficients, from positive to negative.
*$p < .05$.

The estimate of identifying objects, actions, and events and the estimate of deductive reasoning are positive, which suggests that their presence is expected to increase item difficulty. The estimate of flexibility of closure is negative, which suggests that the presence of this feature is expected to decrease item difficulty.

Mathematics

Descriptive results. Twenty-two career-readiness features were used for rating Smarter Balanced items. Of the 186 math items, 65 items were from Grade 8 and 121 items were from Grade 11. For analysis, 145 of the items were from the 2015 test administration and 158 items were from the 2016 test administration. Feature ratings were recoded into binary codes, so that all ratings of 1 and 2 were recoded as 0, and all ratings of 3 and 4 were recoded as 1. Based on this recoded data, math items contained between two and 15 career-readiness features, with an average of 7.75. Table 8 shows the means and standard deviations of the original 4-point ratings. Table 8 also shows the frequency and percentage of items coded as 1 (present) for each career readiness feature.

Table 8

Career-Readiness Features for Math Items (186 Items)

		Original ratings (scale of 1 to 4)		Recoded ratings (0 or 1)	
Related area	Feature	Mean	*SD*	Frequency	Percentage
Skills	Reading comprehension	3.0	0.7	142	76.3
Skills	Complex problem solving	2.0	0.7	47	25.3
Skills	Critical thinking	1.9	0.7	22	11.8
Skills	Active learning	1.1	0.4	9	4.8
Abilities	Deductive reasoning	3.6	0.5	181	97.3
Abilities	Number facility	3.0	1.2	135	72.6
Abilities	Mathematical reasoning	2.6	1.0	108	58.1
Abilities	Inductive reasoning	2.2	0.9	66	35.5
Abilities	Visualization	1.3	0.7	23	12.4
Abilities	Time sharing	1.6	0.7	11	5.9
Abilities	Flexibility of closure	1.4	0.6	6	3.2
Abilities	Information ordering	1.1	0.4	5	2.7
Abilities	Written expression	1.0	0.3	2	1.1
Abilities	Memorization	1.9	0.3	0	0.0
Work activities/context	Analyzing data or information	3.0	0.7	144	77.4
Work activities/context	Importance of being exact or accurate	2.8	0.9	133	71.5
Work activities/context	Processing information	2.7	0.5	126	67.7
Work activities/context	Identifying objects, action, and events	2.5	0.6	90	48.4
Work activities/context	Organizing, planning, and prioritizing work	2.2	0.8	64	34.4
Work activities/context	Getting information	2.1	0.7	48	25.8
Work activities/context	Making decisions and solving problems	2.0	0.8	41	22.0
Work activities/context	Estimating the quantifiable characteristics of products, events, or information	2.1	0.7	39	21.0

Note. Rows are sorted by percentages, in descending order, within each category.

Among the four features related to skills, the more common features rated as present in math items were reading comprehension and complex problem solving. Among the features related to abilities, the more common features were deductive reasoning, number facility, and mathematical reasoning. Among the features related to work activities or context, analyzing data or information and processing information were more common. Features rated as having very little to no presence include active learning, flexibility of closure, information ordering, written expression, and memorization. Several features had an average rating of 2.0 or lower. Unlike the features for ELA items, no features for math items had an average rating of 4.0, and only one feature had an average rating above 3.0 (deductive reasoning). The pattern of ratings was similar across both Grade 8 and Grade 11.

Table 9 shows the total number of features rated as present in math items by feature category.

Table 9

Total Number of Features Rated as Present in Math Items by Feature Category

Number of features	Features related to skills		Features related to abilities		Features related to work activities/context	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
0	40	21.5	–	–	1	0.5
1	91	48.9	25	13.4	13	7.0
2	45	24.2	49	26.3	36	19.4
3	10	5.4	47	25.3	36	19.4
4	–	–	52	28.0	44	23.7
5	–	–	13	7.0	29	15.6
6	–	–	–	–	20	10.8
7	–	–	–	–	5	2.7
8	–	–	–	–	2	1.1
Total	186	100.0	186	100.0	186	100.0

Table 10 shows the correlation matrix of the rating of each feature for math items. The matrix shows the extent to which a feature rated as present was correlated with other features rated as present. Note that memorization was not included due to lack of variability (memorization was rated as not present in any item). In contrast with ELA items, there were no features with high correlations.

Table 10

Correlation Matrix of Features Rated as Present in Math Items

Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. Active learning	–																			
2. Complex problem solving	.10	–																		
3. Critical thinking	**.38**	**.17**	–																	
4. Reading comprehension	.13	**.27**	.13	–																
5. Deductive reasoning	.04	-.13	.06	-.09	–															
6. Flexibility of closure	-.04	-.04	.03	.10	.03	–														
7. Inductive reasoning	-.12	**.40**	-.06	**.31**	-.09	.06	–													
8. Information ordering	-.04	-.10	-.06	.01	.03	-.03	.09	–												
9. Math reasoning	-.06	**.27**	.04	**.35**	.13	.16	**.49**	.07	–											
10. Number facility	.03	-.03	.04	-.06	.12	-.09	.05	.10	**.23**	–										
11. Time sharing	**.48**	.12	**.33**	.09	.04	.21	-.04	-.04	.03	-.05	–									
12. Written expression	-.02	**.18**	**.28**	.06	.02	-.02	.03	-.02	-.02	.06	-.03	–								
13. Visualization	-.09	.08	**.27**	**.17**	-.04	-.07	.13	-.06	-.01	-.21	.04	-.04	–							
14. Analyzing data and information	.12	**.20**	.08	**.40**	-.09	.10	.13	-.07	**.19**	-.16	.14	-.07	.12	–						
15. Estimating	-.12	.10	.10	.19	-.16	-.09	.20	.08	.06	-.10	-.07	.07	**.29**	.06	–					
16. Getting information	**.33**	-.06	**.28**	.07	-.13	**.24**	**-.21**	.05	**-.20**	**-.30**	**.43**	-.06	.00	**.29**	.09	–				
17. Identifying objects, actions, and events	.08	.08	**.25**	.11	-.11	**-.18**	.07	.04	-.03	**-.18**	.08	.11	**.32**	**.16**	**.53**	**.22**	–			
18. Importance of being exact	.03	-.07	.05	-.04	.12	-.09	.02	.03	.21	**.60**	-.04	-.05	-.12	-.14	-.11	**-.28**	-.13	–		
19. Making decisions and solving problems	.12	**.20**	**.29**	**.30**	**-.15**	-.02	**.17**	-.01	.14	.01	**.20**	**.20**	**.27**	**.16**	.04	-.02	**.16**	.02	–	
20. Organizing, planning, and prioritizing work	**.15**	**.26**	**.37**	**.22**	**-.16**	.12	.05	.02	.07	.04	**.20**	.03	**.17**	.09	.13	.14	**.32**	.13	**.16**	–
21. Processing information	.05	-.05	.11	-.09	.10	.00	-.04	.11	**.21**	**.68**	.03	.07	-.13	-.07	-.07	**-.20**	-.05	**.56**	.03	.14

Note. Bold indicates significance at $p < .05$.

Relationship between career-readiness features and item difficulty for mathematics. We employed the same multiple regression method for the career-readiness features in math items as in the ELA items described earlier. Table 11 and Table 12 show the estimates, estimation errors, and *p* values from the analyses of Grade 8 math items and Grade 11 math items, respectively.

Table 11

Relationships Between Item Difficulty Estimate and Career-Readiness Features: Results From Weighted Least Squares Regression for Grade 8 Math

Variable	Estimate	*SE*	*t*	*p*
Intercept	0.806	0.790	1.02	.313
Reading comprehension*	1.829	0.715	2.56	.014
Getting information*	1.550	0.706	2.19	.033
Processing information*	0.883	0.410	2.15	.037
Making decisions and solving problems*	0.791	0.197	4.02	.000
Visualization	0.379	0.578	0.66	.516
Estimating	0.327	0.296	1.11	.275
Number facility	0.218	0.504	0.43	.667
Importance of being exact or accurate	0.131	0.711	0.18	.854
Organizing, planning, and prioritizing work	0.131	0.313	0.42	.678
Inductive reasoning	0.074	0.194	0.38	.706
Math reasoning	-0.123	0.321	-0.38	.705
Identifying objects, actions, and events	-0.171	0.262	-0.65	.518
Complex problem solving	-0.302	0.286	-1.06	.297
Critical thinking	-0.766	0.698	-1.10	.278
Time sharing	-0.816	1.157	-0.71	.484
Information ordering*	-1.775	0.731	-2.43	.019
Analyzing data or information*	-1.906	0.646	-2.95	.005

Note. R square = .53. Features are sorted by coefficients, from positive to negative.
*$p < .05$.

Positive statistically significant coefficients include reading comprehension, getting information, processing information, and making decisions and solving problems, which suggests that these features are associated with an increase in item difficulty in Grade 8 math. The statistically significant negative coefficients of information ordering and analyzing data or information suggest that the presence of these features leads to a decrease in item difficulty.

Table 12

Relationships Between Item Difficulty Estimate and Career-Readiness Features: Results From Weighted Least Squares Regression for Grade 11 Math

Variable	Estimate	*SE*	*t*	*p*
Intercept	1.424	0.317	4.49	<.0001
Written expression	1.322	1.017	1.30	.197
Time sharing*	0.718	0.346	2.07	.041
Math reasoning*	0.624	0.192	3.25	.002
Information ordering	0.452	0.961	0.47	.639
Organizing, planning, and prioritizing work	0.364	0.210	1.74	.086
Reading comprehension	0.323	0.250	1.29	.200
Complex problem solving	0.300	0.218	1.38	.172
Number facility	0.285	0.213	1.34	.183
Importance of being exact or accurate	0.149	0.228	0.65	.515
Identifying objects, actions, and events	-0.016	0.235	-0.07	.945
Visualization	-0.028	0.324	-0.09	.930
Critical thinking	-0.050	0.347	-0.14	.886
Flexibility of closure	-0.140	0.501	-0.28	.781
Inductive reasoning	-0.157	0.202	-0.78	.439
Analyzing data or information	-0.328	0.264	-1.24	.217
Estimating	-0.330	0.295	-1.12	.266
Getting information	-0.384	0.241	-1.60	.113
Processing information	-0.390	0.238	-1.64	.105
Making decisions and solving problems*	-0.681	0.264	-2.58	.011

Note. R square = .45. Features are sorted by coefficients, from positive to negative.
*$p < .05$.

For Grade 11 math, the positive coefficients for time sharing and math reasoning were statistically significant, which suggests that these features are associated with an increase in item difficulty. Making decisions and solving problems was associated with a decrease in item difficulty, in contrast with the Grade 8 results.

Discussion

This study used feature analysis to examine career-readiness features in high school assessments, with a broader goal of refining a set of career-readiness features for standardized

assessments. Results of this study suggest that certain career-readiness features can be found in ELA and math test items. These features were found despite limitations in test item type (largely, multiple choice, multiselect, or short answer). The most frequent features for ELA were importance of being exact or accurate, written comprehension, time sharing, deductive reasoning, and reading comprehension. Some features such as complex problem solving and critical thinking were not found in ELA items, possibly because multiple-choice type items do not lend themselves to complex problem solving or critical thinking. These features were more likely to be rated in the math items. The most frequent features found in math items were deductive reasoning, analyzing data or information, reading comprehension, number facility, and processing information.

This study also suggests that feature rating schemes can be developed and employed by trained raters. The process of applying career-related definitions of skills and abilities to content-based test items was sometimes challenging, and led to few or zero ratings for some features. For instance, ELA items were not rated for active learning, monitoring, problem sensitivity, or visualization, even though these are important skills and abilities for careers. This is not to say that these items contain zero of these skills and abilities, but that they did not contain *enough* in the context of career-based definitions. However, other features were frequently rated and rated highly.

Some differences across grade levels and across content areas were noted in the regression results exploring relationships between the features and expected item difficulty. For instance, reading comprehension was much more strongly associated with item difficulty in Grade 8 math than in Grade 11 math. This may be due to an increase in reading comprehension ability for older students (or, at least, an increase in the comprehension of math items). Future studies might explore such differences through qualitative research. Reading comprehension for Grade 8 ELA, however, was negatively associated with item difficulty. This may suggest a disconnect between student ability and the content of ELA items for Grade 8. Other features, such as flexibility of closure for ELA items and getting information, processing information, and making decisions and solving problems for math items, moved from positive to negative from Grade 8 to Grade 11, which suggests a possible increase in these skills as students progress.

While the findings from this study provide useful information about the relationships between item difficulty and career-readiness features found within test items, examining student test performance data would provide additional insight into career readiness. However, this was not possible due to the nature of computer-adaptive testing. Future work could target specific items with ample student performance data, or explore other sets of tests without this issue. The present study was limited to items from Smarter Balanced's summative, end-of-year assessments. Future work should consider rating and analyzing other types of test items, such as the Smarter Balanced performance tasks, which are geared toward extended problem solving and critical thinking. Such tasks may yield different results and shed additional light on career readiness.

Results from this study suggest that career-readiness features beyond reading comprehension and math are prevalent in summative ELA and math test items. This suggests that inferences about students' career readiness may be drawn from their test scores. Additionally, recognizing and considering the presence of such features can help inform instruction. Helping students prepare for such assessments can be part of the process of preparing students to be career ready, as such skills and abilities can be strengthened with increased practice.

References

Abedi, J., Bayley, R., Ewers, N., Mundhenk, K., Leon, S., Kao, J., & Herman, J. (2012). Accessible reading assessments for students with disabilities. *International Journal of Disability, Development and Education*, *59*(1), 81-95.

Abedi, J., Leon, S., Kao, J., Bayley, R., Ewers, N., Herman, J., & Mundhenk, K. (2011). *Accessible reading assessments for students with disabilities: The role of cognitive, grammatical, lexical, and textual/visual features* (CRESST Report 785). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Baker, E. L., Madni, A., Michiuye, J. K., Choi, K., & Cai, L. (2015). *Mathematical Reasoning Project quantitative analyses results: Grades 4, 8, and 11.* Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Baker, E. L., & O'Neil, H. F., Jr. (2002). Measuring problem solving in computer environments: Current and future states. *Computers in Human Behavior, 18,* 609-622.

Bureau of Labor Statistics. (2017). *Bureau of Labor Statistics*. Washington, DC: U.S. Department of Labor. Retrieved from www.bls.gov

Bureau of Labor Statistics. (2018, April). *College enrollment and work activity of high school graduates news release* (Economic News Release). Retrieved from https://www.bls.gov/news.release/hsgec.htm

Carnevale, A. P., Smith, N., & Strohl, J. (2013). *Recovery: Job growth and education requirements through 2020.* Washington, DC: Georgetown University, Georgetown Public Policy Institute, Center on Education and the Workforce. Retrieved from https://repository.library.georgetown.edu/bitstream/handle/10822/559311/Recovery2020.FR.Web.pdf

Chung, G. K. W. K., & Parks, C. (2015). *Feature analysis validity report* (Deliverable to PBS KIDS). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Chung, G. K. W. K., & Redman, E. H. (2015a). *Feature analysis framework – final* (Deliverable to PBS KIDS). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Chung, G. K. W. K., & Redman, E. H. (2015b). *Feature analysis ratings best practices* (Deliverable to PBS KIDS). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

EDFacts. (2018). *Data files*. Washington, DC: U.S. Department of Education. Retrieved from http://www.ed.gov/edfacts

Gordon, E. W. (1970). Toward a qualitative approach to assessment. *Report of the Commission on Tests, II. Briefs* (pp. 42-46). New York: College Entrance Examination Board.

Gordon, E. W., & Rajagopalan, K. (2016). Assessment for teaching and learning, Not just accountability. In E. W. Gordon & K. Rajagopalan (Eds.), *The testing and learning revolution: The future of assessment in education* (pp. 9-34). New York, NY: Palgrave Macmillan.

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology: Research & Development, 48*(4), 63-85.

Madni, A., Kao, J. C., Rivera, N. M., Baker, E. L., & Cai, L. (2018). *Exploring career-readiness features in high school test items through cognitive laboratory interviews* (CRESST Report 857). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Mejia, M. C., Rodriguez, O., & Johnson, H. (2016). *Preparing students for success in California's community colleges.* Retrieved from http://www.ppic.org/content/pubs/report/R_1116MMR.pdf

Roberts, J. D., Chung, G. K. W. K., & Parks, C. (2016). Supporting children's progress through the PBS KIDS learning analytics platform. *Journal of Children and Media, 10*(2). Retrieved from http://dx.doi.org/10.1080/17482798.2016.1140489

Smarter Balanced. (2017, September 18). *South Dakota to use Smarter Balanced scores for college admissions*. Retrieved from http://www.smarterbalanced.org/south-dakota-use-smarter-balanced-scores-college-admissions/

Wolf, M. K., Herman, J. L., Kim, J., Abedi, J., Leon, S., Griffin, N., ... Shin, H. W. (2008). *Providing validity evidence to improve the assessment of English language learners* (CRESST Report 738). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Wolf, M. K., Kim, J., Kao, J. C., & Rivera, N. M. (2009). *Examining the effectiveness and validity of glossary and read-aloud accommodations for English language learners in a math assessment* (CRESST Report 766). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

UCLA

CRESST

NATIONAL CENTER FOR RESEARCH ON EVALUATION, STANDARDS, AND STUDENT TESTING

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)

Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522

(310) 206-1532
www.cresst.org