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

# CAREER-READINESS FEATURES IN KOREAN ASSESSMENT ITEMS

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

DECEMBER 2018 CRESST REPORT **859**

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).



To cite from this report, please use the following as your APA reference: Choi, K., Kao, J. C., Rivera, N. M., & Cai, L. (2018). *Career-readiness features in Korean assessment items* (CRESST Report 859). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

## Table of Contents

# Career-Readiness Features in Korean Assessment Items[1]

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

CRESST/University of California, Los Angeles

**Abstract:** This report is the third in a series considering career-readiness features within high school assessments. The goal of this study was to explore international comparisons by applying feature analysis to Korean assessment items. Twenty math test items from the Gyeonggi Province in South Korea along with performance data from roughly 4,000 Grade 12 examinees were obtained. We applied the process of feature rating and analyses explained in the first two reports in this series to the Korean assessment items. Math test items were translated and rated using the same scheme as the American items by the same math expert raters. While the mathematics content differed from the American test, career-readiness features were also found in this small set of items. Korean math items shared similar patterns of feature representation with Smarter Balanced items, with the exception of a couple of features (e.g., critical thinking and visualization were rated more frequently among Korean items). Critical thinking was positively associated with item difficulty. It is possible that the type of math content (e.g., calculus) plays a role, which should be explored in future studies with larger sets of test items, different math content, and other international data sets.

## Introduction

The purpose of this study was to continue CRESST's career-readiness feature analysis work by exploring international comparisons with assessment data from The Republic of Korea (hereafter, Korea). Korea is widely recognized for achievement in primary and secondary education, particularly in international assessments (So & Kang, 2014). For example, Korea regularly outpaces the United States and the international average on assessments such as PISA (Programme for International Student Assessment) in reading, science, and mathematics (Organisation for Economic Co-operation and Development [OECD], 2015). Since this

[1]We thank Eric Zilbert of the California Department of Education for support and guidance. We are grateful to Eva Baker and Noelle Griffin for their advice and support. We also thank Seung Won Chung, Markus Iseli, Eunhee Keum, Junok Kim, Ayesha Madni, Julia Phelan, and Yoon Soo Suh for their contributions to this research. Special thanks to Robert Mislevy and David Thissen for reviewing early drafts of this report.

benchmarking study will focus on assessment metrics, some context about the educational policies and practices in Korea may be useful.

The educational system in Korea has some similarities with education in the United States. The Korean education system follows a "6–3–3–4" model: six years of primary schooling, three years of middle school, three years of high school, and four years of undergraduate education. The Korean government offers both vocational and academic high school tracks, and started offering preschool for children in 2012 (National Center on Education and the Economy [NCEE], 2018). Primary and secondary school are compulsory until the age of 15 (Korean Council for University Education, 2015; Kwon, Lee, & Shin, 2015). The compulsory education system in the U.S. roughly mirrors this pattern but can vary from state to state. Attendance in the U.S. is generally required from about age 6 until at least age 16, with students generally continuing education until they complete Grade 12 (National Center for Education Statistics, n.d.-a).

As the U.S. struggles with educational reforms and the balance between learning, assessment, and career readiness, so too does Korea. Korean schools follow a national curriculum framework developed by the Ministry of Education (n.d.-a), which was revised in 2015. The Korean government also passed the Career Education Act in 2015, to clarify the responsibilities in providing effective career education for students. The newer curriculum is aimed at reducing content and increasing choice in schools (So & Kang, 2014). Approximately 20% of students enrolled in upper secondary schools are on vocational tracks, down from 40% in the late 1990s (NCEE, 2018). The curriculum which previously diverged into vocational and academic tracks has been merged into a single ladder pathway. To increase career awareness, Korean students are now required to spend at least one hour per week in career exploration beginning in lower secondary school (NCEE, 2018) and have "free learning semesters" aimed at career exploration beginning in middle school (International Centre for Career Development and Public Policy, 2017).

Beyond secondary (high school) education, both the United States and Korea have higher than average numbers of 25- to 64-year-olds with tertiary (or postsecondary) education compared to other OECD (Organisation for Economic Co-operation and Development) member countries (46% and 47%, respectively; OECD, 2018). However, academic achievement may come at a cost. As of 2010, a reported 74% of Korean students engaged in private tutoring such as *hagwons*, or "cram schools," leading to government interventions to reduce student stress (Ripley, 2011). Compared to youth in 30 other OECD countries, Korean youth spend an average of three more hours per day studying ("Korean youth study longest hours in OECD," 2009). They sleep one less hour and exercise 22 minutes less than youth in five other OECD countries, including the United States. Beyond high school, the Korean Ministry of Education reports that about 71% of students go on to pursue tertiary education (Ministry of Education, n.d.-b). The emphasis on academic performance has led to unintended consequences such as highly stressed students, lower student interest in learning, increasingly deskilled teachers because of the prescribed national curriculum (So & Kang, 2014), and "excessive" college enrollment for available jobs (International Centre for Career Development and Public Policy, 2017).

Similar to the U.S., Korea has a system of assessments that are administered nationally, by province (district in the U.S.) or at the school level. In California, students in Grades 3–8 and 11 take the California Assessment of Student Performance and Progress (CASPP) including math and ELA (National Center for Education Statistics, n.d.-b). Korea has a system of assessments called the National Assessment of Educational Achievement (NAEA). Each year, the NAEA is administered to all students in Grades 9 and 11, in the subjects of Korean, math, and English. Tests in science and social studies are administered to a sample of students in Grade 9. Results are reported in aggregate and used to provide additional support for schools and inform policy (NCEE, 2018).

The purpose of the present study was to provide a preliminary international benchmark, specifically, comparing the career-readiness features found in the Gyeonggi Province Mathematics Test with the career-readiness features found in the Grade 11 Smarter Balanced mathematics test administered in the state of California, which were reported in a companion report (Kao, Choi, Rivera, Madni, & Cai, 2018). The Korean test items used in this study were translated from a 12$^{th}$ grade mathematics test developed in alignment with national standards in order to evaluate educational accountability in the Gyeonggi Province (Gyeonggi Institute of Education, 2014). We applied the process of feature rating and analyses developed in our prior report to the Korean test items. As discussed in Kao et al. (2018), career-readiness features were found within high school assessment items, which suggests that inferences can be drawn about students' career readiness from their test scores. We were interested in whether career-readiness features were also found in Korean items, and if so, which features. We were also interested in which particular features had relationships with item difficulty. Results could pave the pathway for future international benchmarking studies.

# Method

## Test Items

For this study, we examined a 12th-grade-level mathematics achievement test from Korea, which was administered in 2014 by the Gyeonggi Institute of Education. The Gyeonggi Institute of Education is a Korean-government-funded educational research institute. This test consists of 20 multiple-choice items covering calculus concepts, including limits of a function, continuous functions, differentiation, higher order derivatives, application of differentiation and derivatives, indefinite integrals, definite integrals, and applications of indefinite and definite integrals. In addition, according to the test blueprint, test task domains include concepts and procedures, problem solving, communication, and inferences (Gyeonggi Institute of Education, 2014).

The items were independently translated into English by three native Korean-speaking bilingual graduate students, with experience in both the Korean and United States educational systems. These translators were familiar with the math content and did not require initial content training. The translated items were then translated from English back into Korean ("back-translated"), as described by Brislin (1970). The translations were validated by a fourth native Korean speaker as free from errors due to literal translations or "meaning errors" (Brislin, 1970, p. 197). Translations were slightly edited for grammar and style. See the Appendix for the translated items.

## Target Career-Readiness Features

**Features for math test items.** As described in our prior report on feature analysis, following subject-matter expert review, an initial list of 36 career-readiness features was created, and then narrowed by math content area experts into the final list of 22 features (Kao et al., 2018). Table 1 shows the list of 22 career-readiness features used to rate math items, with descriptions. They are grouped broadly into three categories: features related to skills, features related to abilities, and features related to either work activities or work context.

Table 1

*Career-Readiness Target Features Used to Rate Math Test Items 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. |
| 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. |
| 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. |
| 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. |
| Time sharing | The ability to shift back and forth between two or more activities or sources of information. |
| 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. |
| 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. |

| Feature | Description |
|---|---|
| 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. |
| 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 rating.** The three raters who previously rated Smarter Balanced math items, described in Kao et al. (2018), used the same procedure and feature list to rate the Korean test items. As described in the prior report, the math raters held doctorates or master's degrees in fields related to engineering, measurement, and statistics and were familiar with the content. Test items were rated by all three raters using an ordinal 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. The Krippendorff's alpha for interrater reliability ratings was .86 (Krippendorff, 2011). Raters discussed disagreements to achieve final consensus. As in the analyses on the Smarter Balanced items, ratings were subsequently recoded from ordinal values into binary values as the focus was on the frequency or presence of a feature. Ratings of 1 and 2 were recoded as 0, indicating the feature was not present, and ratings of 3 and 4 were recoded as 1, indicating the feature was present. The binary coding was used with item difficulty analyses.

# Results

## Descriptive Results

We present descriptive results from the Grade 11 Smarter Balanced feature ratings alongside results from the Korean item feature ratings. Table 2 shows the means and standard deviations of career-readiness features using the original 4-point rating scale for Smarter Balanced and Korean math items. The range of standard deviations were similar for each set of test items. Based on descriptive results, Korean math items were about as high as Smarter Balanced math items on features such as reading comprehension, deductive reasoning, mathematical reasoning, number facility, analyzing data or information, getting information, importance of being exact or accurate, and processing information. They were also similarly low in features such as active learning, flexibility of closure, information ordering, memorization, time sharing, and written expression.

Table 2

*Means and Standard Deviations of Career-Readiness Features Using Original 4-Point Rating Scale, Smarter Balanced vs. Korean Math Items*

| | Smarter Balanced (121 items) | | Korean (20 items) | |
|---|---|---|---|---|
| Feature | *M* | *SD* | *M* | *SD* |
| **Features related to skills** | | | | |
| Reading comprehension | 3.1 | 0.7 | 2.8 | 0.8 |
| Complex problem solving | 2.1 | 0.7 | 2.0 | 0.0 |
| Critical thinking | 2.0 | 0.8 | 2.5 | 1.2 |
| Active learning | 1.2 | 0.5 | 1.3 | 0.9 |
| **Features related to abilities** | | | | |
| Deductive reasoning | 3.6 | 0.6 | 3.0 | 0.0 |
| Number facility | 2.8 | 1.2 | 3.6 | 0.9 |
| Mathematical reasoning | 2.5 | 0.9 | 2.4 | 0.9 |
| Inductive reasoning | 2.2 | 0.8 | 1.1 | 0.2 |
| Memorization | 1.8 | 0.4 | 1.1 | 0.2 |
| Time sharing | 1.7 | 0.7 | 1.4 | 0.6 |
| Flexibility of closure | 1.4 | 0.6 | 1.2 | 0.4 |
| Visualization | 1.3 | 0.7 | 1.6 | 0.9 |
| Information ordering | 1.1 | 0.3 | 1.0 | 0.0 |
| Written expression | 1.1 | 0.4 | 1.0 | 0.0 |
| **Features related to work activities/context** | | | | |
| Analyzing data or information | 3.1 | 0.7 | 3.4 | 0.5 |
| Processing information | 2.6 | 0.5 | 3.0 | 0.0 |
| Identifying objects, action, and events | 2.6 | 0.6 | 2.1 | 0.3 |
| Importance of being exact or accurate | 2.5 | 0.8 | 3.0 | 0.8 |
| Organizing, planning, and prioritizing work | 2.3 | 0.8 | 2.6 | 0.5 |
| Getting information | 2.2 | 0.8 | 2.5 | 0.8 |
| Making decisions and solving problems | 2.0 | 0.6 | 2.0 | 0.0 |
| Estimating the quantifiable characteristics of products, events, or information | 2.0 | 0.7 | 1.1 | 0.4 |

*Note*. Features are sorted in descending order of Smarter Balanced means within each category.

Table 3 shows the frequency and percentage of each career-readiness feature using the binary ratings for both Smarter Balanced and Korean math items. Under the binary coding, the 20 Korean math items contained between five and 12 career-readiness features, with a mean of 7.95. In comparison, the 121 Smarter Balanced math items contained between two and 15 features, with a mean of 7.77. Some features were rated more frequently among Smarter Balanced items (such as complex problem solving and inductive reasoning), while some features were rated more frequently among Korean items (such as critical thinking and visualization).

Table 3

*Frequencies of Career-Readiness Features Using Recoded Binary Ratings (Smarter Balanced vs. Korean Math Items)*

| | Smarter Balanced (121 items) | | Korean (20 items) | |
|---|---|---|---|---|
| Feature | Frequency | % | Frequency | % |
| **Features related to skills** | | | | |
| Reading comprehension | 97 | 80.2 | 12 | 60.0 |
| Complex problem solving | 37 | 30.6 | 0 | 0.0 |
| Critical thinking | 17 | 14.0 | 10 | 50.0 |
| Active learning | 8 | 6.6 | 2 | 10.0 |
| **Features related to abilities** | | | | |
| Deductive reasoning | 116 | 95.9 | 20 | 100.0 |
| Number facility | 81 | 66.9 | 18 | 90.0 |
| Mathematical reasoning | 67 | 55.4 | 11 | 55.0 |
| Inductive reasoning | 43 | 35.5 | 0 | 0.0 |
| Visualization | 15 | 12.4 | 6 | 30.0 |
| Time sharing | 10 | 8.3 | 1 | 5.0 |
| Flexibility of closure | 6 | 5.0 | 0 | 0.0 |
| Information ordering | 2 | 1.7 | 0 | 0.0 |
| Written expression | 2 | 1.7 | 0 | 0.0 |
| Memorization | 0 | 0.0 | 0 | 0.0 |
| **Features related to work activities/context** | | | | |
| Analyzing data or information | 98 | 81.0 | 20 | 100.0 |
| Importance of being exact or accurate | 72 | 59.5 | 18 | 90.0 |
| Processing information | 72 | 59.5 | 20 | 100.0 |
| Identifying objects, action, and events | 62 | 51.2 | 2 | 10.0 |
| Organizing, planning, and prioritizing work | 48 | 39.7 | 12 | 60.0 |
| Getting information | 41 | 33.9 | 6 | 30.0 |
| Estimating the quantifiable characteristics of products, events, or information | 25 | 20.7 | 1 | 5.0 |
| Making decisions and solving problems | 21 | 17.4 | 0 | 0.0 |

Table 4 shows the total number of features rated as present in Korean math items by feature category.

Table 4

*Frequency and Percentage of Features Rated as Present in Korean Math Items by Feature Category*

| Number of features | Features related to skills | | Features related to abilities | | Features related to work activities/context | |
|---|---|---|---|---|---|---|
| | Frequency | % | Frequency | % | Frequency | % |
| 0 | 5 | 25.0 | - | - | - | - |
| 1 | 8 | 40.0 | 1 | 5.0 | - | - |
| 2 | 7 | 35.0 | 6 | 30.0 | - | - |
| 3 | - | - | 9 | 45.0 | 8 | 40.0 |
| 4 | - | - | 4 | 20.0 | 6 | 30.0 |
| 5 | - | - | - | - | 5 | 25.0 |
| 6 | - | - | - | - | 1 | 5.0 |
| Total | 20 | 100.0 | 20 | 100.0 | 20 | 100.0 |

Table 5 shows the career-readiness features rated by test item number. Each item contained between five and 12 features, with a total of 15 career-readiness features represented across the test items.

Table 5

*Career-Readiness Features by Test Item*

| | Test item number | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Feature | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | Total |
| Deductive reasoning | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | 20 |
| Analyzing data or information | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | 20 |
| Processing information | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | 20 |
| Number facility | | X | X | X | X | X | X | X | X | X | X | X | X | | X | X | X | X | X | X | 18 |
| Importance of being exact or accurate | X | X | X | X | X | | X | X | X | X | X | X | X | | X | X | X | X | X | X | 18 |
| Reading comprehension | | | X | X | X | X | | | | X | X | | X | X | X | X | | X | X | | 12 |
| Organizing, planning, and prioritizing work | | X | X | | X | X | | | X | X | X | X | X | X | | | | X | X | | 12 |
| Mathematical reasoning | X | X | X | X | X | X | X | | | X | | | | | | X | X | | | X | 11 |
| Critical thinking | | X | X | | X | X | | | X | X | X | X | X | | | X | | | | | 10 |
| Visualization | | | | | X | X | | | X | | X | X | | | | | | | | X | 6 |
| Getting information | | | X | | X | | | | | X | | | X | X | | | | | X | | 6 |
| Active learning | | | | | | | | | | X | | | | | | | | | X | | 2 |
| Identifying objects, actions, and events | | | | | | X | | | | | | | | X | | | | | | | 2 |
| Time sharing | | | | | | | | | | X | | | | | | | | | | | 1 |
| Estimating the quantifiable characteristics of products, events, or information | | | | | X | | | | | | | | | | | | | | | | 1 |

| Feature | Test item number | | | | | | | | | | | | | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | |
| Complex problem solving | | | | | | | | | | | | | | | | | | | | | 0 |
| Flexibility of closure | | | | | | | | | | | | | | | | | | | | | 0 |
| Inductive reasoning | | | | | | | | | | | | | | | | | | | | | 0 |
| Information ordering | | | | | | | | | | | | | | | | | | | | | 0 |
| Memorization | | | | | | | | | | | | | | | | | | | | | 0 |
| Written expression | | | | | | | | | | | | | | | | | | | | | 0 |
| Making decisions and solving problems | | | | | | | | | | | | | | | | | | | | | 0 |
| Total | 5 | 8 | 10 | 7 | 12 | 10 | 6 | 5 | 8 | 12 | 9 | 8 | 9 | 7 | 6 | 8 | 6 | 7 | 9 | 7 | |

Table 6 shows a correlation matrix of the feature rating for Korean math items. The matrix shows the extent to which a feature rated as present was correlated with other features rated as present. For example, if critical thinking was rated as present in a Korean item, such an item was likely to also be rated for visualization as well as organizing, planning, and prioritizing work. Note that features without any variability were *not* included. That is, deductive reasoning, analyzing data or information, and processing information were rated as present across all 20 items and were excluded. Similarly, complex problem solving, flexibility of closure, inductive reasoning, information ordering, memorization, written expression, and making decisions and solving problems were not rated as present in any of the items and were also excluded.

Unlike Smarter Balanced math items, some features showed strong correlations. For instance, the importance of being exact or accurate showed a negative 1.0 correlation with identifying objects, actions, or events. However, because the correlations are based on only 20 items with binary values (present or not present), the correlations should be interpreted with some caution.

Table 6

*Correlation Matrix of Features Rated as Present in Korean Math Items*

| Feature | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Active learning | — | | | | | | | | | | |
| 2. Critical thinking | .00 | — | | | | | | | | | |
| 3. Reading comprehension | .27 | .20 | — | | | | | | | | |
| 4. Mathematical reasoning | -.03 | .10 | -.12 | — | | | | | | | |
| 5. Number facility | .11 | .33 | .07 | .03 | — | | | | | | |
| 6. Time sharing | **.69**** | .23 | .19 | .21 | .08 | — | | | | | |
| 7. Visualization | -.22 | .44 | -.13 | -.07 | .22 | -.15 | — | | | | |
| 8. Estimating | -.08 | .23 | .19 | .21 | .08 | -.05 | .35 | — | | | |
| 9. Getting information | **.51*** | .22 | **.54*** | -.07 | -.15 | .35 | -.19 | .35 | — | | |
| 10. Identifying, objects, actions, events | -.11 | .00 | .27 | -.03 | **-.44*** | -.08 | .15 | **.69**** | **.51*** | — | |
| 11. Importance of being exact or accurate | .11 | .00 | -.27 | .03 | **.44*** | .08 | -.15 | **-.69**** | **-.51*** | **-1.00**** | — |
| 12. Organizing, planning, and prioritizing work | .27 | **.61**** | .38 | -.33 | .07 | .19 | .31 | .19 | **.54*** | .27 | -.27 |

**p* < .05, two-tailed. ***p* < .01, two-tailed.

## Relationship Between Career-Readiness Features and Item Difficulty

One goal of the present study was to explore whether the career-readiness features within a test item was associated with the item's difficulty. Item response theory (IRT) allows estimates of item difficulty and test-taker ability and is widely used in large-scale educational assessment (Thissen & Steinberg, 2009). We first examined the item difficulty parameter and discrimination parameters using the two-parameter item response model (2PL). Data were obtained from the 2014 Gyeonggi Institute of Education test administration, and included 3,949 records examinee records.

Item responses were coded as 0 for incorrect or 1 for correct. Descriptive statistics of the item difficulty parameters were computed. The mean item difficulty was 0.49 ($SD$ = 0.69), with a minimum of -0.78 and a maximum of 1.65. Table 7 shows the distribution of item difficulty in the Korean items. As shown in the figure, more than three quarters of the items have positive values of item difficulty. However, as the minimum and maximum values of the item difficulty indicate, there were no items that were extremely easy or extremely difficult.

Table 7

*Item Difficulty in Korean Items*

| Item | Item difficulty (IRT-B) | *SE* |
|---|---|---|
| 1 | -0.78 | 0.04 |
| 2 | 0.16 | 0.04 |
| 3 | 0.35 | 0.03 |
| 4 | -0.45 | 0.03 |
| 5 | 0.54 | 0.04 |
| 6 | 1.43 | 0.08 |
| 7 | -0.32 | 0.03 |
| 8 | -0.26 | 0.03 |
| 9 | 1.49 | 0.14 |
| 10 | 0.75 | 0.05 |
| 11 | 0.87 | 0.07 |
| 12 | 0.72 | 0.06 |
| 13 | 1.30 | 0.09 |
| 14 | 0.89 | 0.05 |
| 15 | 0.18 | -0.03 |
| 16 | 1.65 | 0.10 |
| 17 | -0.02 | 0.03 |
| 18 | 0.71 | 0.05 |
| 19 | 0.46 | 0.04 |
| 20 | 0.40 | 0.03 |

A multiple regression technique, imposing precision weights, was employed to examine the relationship between item difficulty and identified career readiness features. Weight was calculated by 1/([Standard Error of item difficulty]$^2$/the total number of examinees). We first explored the relationship using a longer list of career-readiness features as predictors and item difficulty as the outcome. However, with a limited number of items and correlations among predictors, we then chose a reduced feature list—critical thinking, reading comprehension, math reasoning, and visualization. The weighted multiple regression result is presented in Table 8.

The critical thinking feature had a statistically significant positive relationship with item difficulty. In other words, the item difficulty increased by 0.53 if an item rated with the critical thinking feature was compared to an item without it. In addition, an item with the visualization feature is considered more difficult by approximately 0.50 scale in difficulty than an item without the visualization feature. The *R*-square of this model was .56, which is considered substantial given the small number of predictors.

Table 8

*Relationships Between Item Difficulty Estimate and Career-Readiness Features: Results From Weighted Multiple Regression for Korean Math Items*

| Variable | Estimate | *SE* | *t* | *p* |
|---|---|---|---|---|
| Intercept | -0.034 | 0.195 | -0.17 | .866 |
| Critical thinking* | 0.525 | 0.220 | 2.38 | .031 |
| Visualization | 0.497 | 0.233 | 2.14 | .050 |
| Reading comprehension | 0.205 | 0.195 | 1.05 | .310 |
| Math reasoning | -0.294 | 0.198 | -1.49 | .158 |

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

To ensure the appropriateness of model selection, we also employed a weighted multiple regression with stepwise selection method. As shown in Table 9, the results are similar to those in the previous model. The *R*-square of this model was .53. The result suggests that for 12th grade Korean students, items measuring features such as critical thinking or visualization are more difficult than items without such features. Items measuring features such as math reasoning tend to be easier than items without math reasoning present.

Table 9

*Relationships Between Item Difficulty Estimate and Career Readiness Features: Results From Weighted Multiple Regression With Stepwise Selection Method for Korean Math Items*

| Variable | Estimate | *SE* | *t* | *p* |
|---|---|---|---|---|
| Intercept | 0.095 | 0.153 | 0.38 | .544 |
| Critical thinking* | 0.616 | 0.203 | 9.24 | .008 |
| Visualization | 0.451 | 0.229 | 3.88 | .067 |
| Math reasoning | -0.357 | 0.189 | 3.56 | .078 |

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

Results differ from those found in the 11th grade Smarter Balanced math items, in which several features contained coefficient estimates larger than .5 (positive: written expression, time sharing, math reasoning; negative: making decisions and solving problems). Korean math items were all rated as 0 for written expression and for making decisions and solving problems, and only one of the 20 Korean items was rated as having time sharing. It is possible that Korean math items do not contain such features, though other test items from Korea may have yielded different results.

# Discussion

In this report, we explore preliminary international comparisons with a 20-item multiple-choice math test from the Gyeonggi province in Korea. Similar to Smarter Balanced summative math test items (Kao et al., 2018), certain career-readiness features can also be found in Korean math test items. While the mathematics content differs from the American test, career-readiness features were also found in this small set of items. These features were present even though response types were multiple choice for Korean items and varied for Smarter Balanced items. This study represents a preliminary comparison of Korean and Smarter Balanced test items on mathematics and evidence that an international assessment can be rated using the feature rating process.

However, some limitations of the study and differences in our samples require caution in making direct comparisons. For example, in terms of the features in the respective tests, a larger proportion of Korean items contained critical thinking and visualization, which were also positively associated with item difficulty. This *may* be reflective of differences in culture or educational pedagogy. However, this may also reflect differences in content. That is, while Smarter Balanced math item content was more focused on algebraic concepts, the Korean items generally covered calculus. Perhaps calculus requires more critical thinking and visualization than algebra. In informal conversations with the math raters, they remarked that the Korean items generally had more items that involved multiple steps to solve the problem (i.e., critical thinking) and that the Korean items were more likely to require either visualizing or drawing out graphs. Another limitation to this study is that the Korean test items were administered for local accountability and may not reflect the same attributes as summative tests used to gain entry at the university level. Apart from content differences, future research might consider individual differences such as age or motivation.

Overall, this study represents a preliminary comparison of Korean and Smarter Balanced test items on mathematics and evidence that international assessment can be rated using the feature rating process. In light of educational focuses on career readiness in the U.S. and abroad, future research should explore a larger set of test items to explore whether trends continue, and whether there are implications of differing math content on career readiness for high school students. This may include calculus-based assessments and other mathematics content from other countries. Such research can shed additional light on drawing career-readiness inferences from mathematics assessments.

# References

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*(3), 185-216.

Gyeonggi Institute of Education. (2014). *Geospatial training.* Suwon, South Korea: Author.

International Centre for Career Development and Public Policy. (2017). *About career education in Korea.* Retrieved from http://iccdpp2017.org/is-2017-in-korea/about-career-education-in-korea/?ckattempt=1

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.

Korean Council for University Education. (2015). *Korea's educational system.* Retrieved from http://adiga.kr/kcue/ast/eip/eng/EngGnrl.do?p_menu_id=30101#page6

Korean youth study longest hours in OECD. (2009, August 10). *The Chosun Ilbo.* Retrieved from http://english.chosun.com/site/data/html_dir/2009/08/10/2009081000200.html

Krippendorff, K. (2011). *Computing Krippendorff's alpha-reliability*. Retrieved from http://repository.upenn.edu/asc_papers/43

Kwon, S. K., Lee, M., & Shin, D. (2015). Educational assessment in the Republic of Korea: Lights and shadows of high-stake exam-based education system. *Assessment in Education: Principles, Policy & Practice*, *24*(1), 60-77. doi: 10.1080/0969594X.2015.1074540

Ministry of Education. (n.d.-a). *Metropolitan & Provincial Offices of Education*. Retrieved from http://english.moe.go.kr/sub/info.do?m=0803&s=english

Ministry of Education. (n.d.-b). Table: *Rates of advancement to next-level school*. Retrieved from http://english.moe.go.kr/sub/info.do?m=040101&s=english

National Center for Education Statistics. (n.d.-a). *Table 5.1 Compulsory school attendance laws, minimum and maximum age limits for required free education, by state: 2015 from State Education Reforms (SER).* Washington, DC: U.S. Department of Education. Retrieved from https://nces.ed.gov/programs/statereform/tab5_1.asp

National Center for Education Statistics. (n.d.-b). *Table 2.3. English/Language Arts (ELA) and mathematics statewide high school assessments, by state: 2016–17 from State Education Reforms (SER).* Washington, DC: U.S. Department of Education. Retrieved from https://nces.ed.gov/programs/statereform/tab2_3.asp

National Center on Education and the Economy. (2018). *South Korea: Learning Systems.* Retrieved from http://ncee.org/what-we-do/center-on-international-education-benchmarking/top-performing-countries/south-korea-overview/south-korea-instructional-systems/

OECD. (2015). *OECD skills outlook 2015: Youth, skills and employability*. Paris, France: Author. Retrieved from http://dx.doi.org/10.1787/9789264234178-en

OECD. (2018). *Adult education level (indicator).* Retrieved from doi: 10.1787/36bce3fe-en

Ripley, R. (2011, September 25). Teacher, leave those kids alone. *Time.* Retrieved from http://content.time.com/time/magazine/article/0,9171,2094427,00.html

So, K., & Kang, J. (2014). Curriculum reform in Korea: Issues and challenges for twenty-first century learning. *The Asia-Pacific Education Researcher, 23,* 795-803. doi:10.1007/s40299-013-0161-2

Thissen, D., & Steinberg, L. (2009). Item response theory. In R. Millsap & A. Maydeu Olivares (Eds.), *The Sage handbook of quantitative methods in psychology* (pp. 148-177). London, England: Sage. Retrieved from http://dx.doi.org/10.4135/9780857020994.n7

# Appendix:
# Translated Test Items

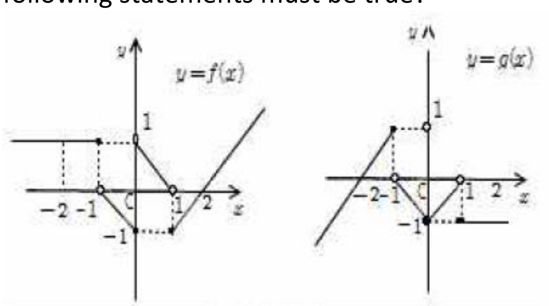
1. Evaluate $\lim_{x\to\infty}\frac{5x^2-9}{2x^2-3}$

① $\frac{2}{5}$
② 1
③ $\frac{9}{5}$
④ $\frac{5}{2}$
⑤ 3

2. For two constants $a$ and $b$, if $\lim_{x\to 1}\frac{\sqrt{x^2+a}-2}{x-1} = b$, find $a + 2b$?

① 3
② 4
③ 5
④ 6
⑤ 7

3. The graphs of the functions $y = f(x)$ and $y = g(x)$ are shown in the figures below. Which of the following statements must be true?

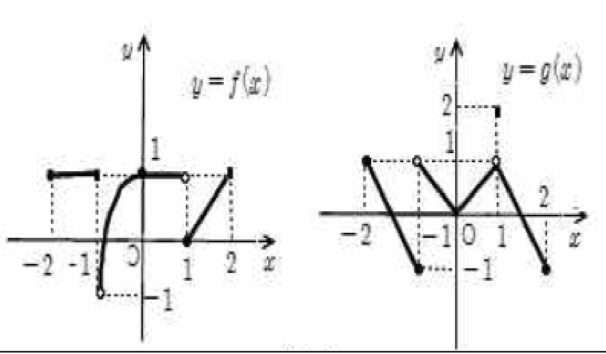



A. $\lim_{x\to 2+0}\frac{g(x)}{f(x)} = 0$
B. $\lim_{x\to -0} f(x)g(x) = 0$
C. $\lim_{x\to 0}\{f(x) - g(x)\} = 2$

① A
② B
③ A, C
④ B, C
⑤ A, B, C

4. A function $f(x)$ is continuous in the set of all real numbers. If $f(x)$ satisfies $(x-2)f(x) = x^2 - 5x + 6)$, find $f(2)$.
① -3
② -1
③ 2
④ 4
⑤ 6

5. The graphs of two functions $y = f(x)$ and $y = g(x)$, defined in the interval $[-2, 2]$, are shown below. Which of the following statements must be true?




A. A function $f(x)g(x)$ is continuous at $x = -1$.
B. A function $(1-x)g(x)$ is continuous at $x = 1$.
C. A function $f(x) + g(x)$ is continuous at $x = 0$.

① A
② C
③ A, B
④ B, C
⑤ A, B, C

6. A function $f(x)$ is continuous in the interval $[1, 2]$ and satisfies $f(1) = 2a^2 - a$ and $f(2) = a - 3$. What is the sum of all possible values of $a$ which makes $f(x) - x = 0$ have at least one real root on the interval $[1, 2]$?
① 9
② 10
③ 12
④ 14
⑤ 17

7. For a differentiable function $f(x)$ with $f'(a) = 5$, what is the value of $\lim_{h\to 0} \frac{f(a+3h)-f(a)}{h}$?
① 6
② 9
③ 12
④ 15
⑤ 18

8. For function $f(x) = (x^2 + 1)(2x - 1)$, find $f'(1)$.
① 2
② 3
③ 4
④ 5
⑤ 6

9. For function $f(x) = 3x^4 + 4ax^3 - 6x^2 + 12ax$, for $x > 1$, what is the range of constant $a$?
① $a > -1$
② $a \leq -1$
③ $a \leq 1$
④ $a > 1$
⑤ $-1 < a \leq 1$

10. A cubic function $f(x) = x^3 + ax^2 + bx + c$ has the following conditions:

A. $f(-x) = -f(x)$

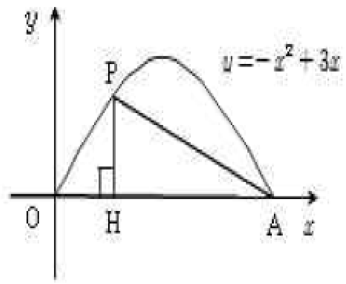
B. At a point on $y = f(x)$, the slope of the tangent line at $x = 2$ is 9.

What is the local maximum of $f(x)$?
① 1
② 2
③ 3
④ 4
⑤ 5

11. The tangent line to the curve $y = x^3 - 2$ at point $(0,0)$ meets the curve at a point of tangency *A* and point *B*. What is the length of $\overline{AB}$?
① $\sqrt{10}$
② $2\sqrt{10}$
③ $3\sqrt{10}$
④ $4\sqrt{10}$
⑤ $5\sqrt{10}$

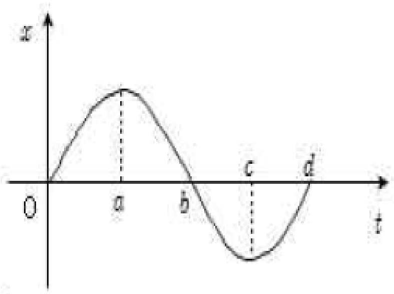
12. When a cubic function $f(x) = 4x^3 + 3x^2$ is translated parallel to the Y-axis by *k* units and meets the X-axis at three different points, what is the range of k?
①$-1 < k < \frac{1}{2}$
②$-\frac{1}{4} < k < 0$
③ $k < -1, k > \frac{1}{2}$
④ $\frac{1}{3} < k < 1$
⑤$k < -\frac{1}{4}, k > 0$

13. As shown in the figure below, let a point on curve $y = -x^2 + 3x$ $(0 < x < 3)$ be $P$ and let $H$ be the point from point P to the X-axis, forming a perpendicular line. Point $A$ is the intersection where the curve intersects with the x-axis other than the origin 0. Suppose the area of $\Delta PAH$ is maximum. What is the length of a segment $AH$?




① $\frac{1}{2}$
② 1
③ $\frac{3}{2}$
④ 2
⑤ $\frac{5}{2}$

14. Suppose a point $P$ moves from the origin through a vertical line. Let $x$ be the location of the point $P$ at time $t$ as shown in the figure, which depicts $x = f(t)$. Which of the following statements below must be true? (Note that $0 \leq t \leq d$)




A. The point $P$ shifts its direction twice while it moves through the line.
B. The velocity of point $P$ when it passes the origin the first time after departure is $f'(b)$.
C. The direction of the movement is identical when $0 < t < a$ and $0 < t < d$.

① A
② B
③ A, C
④ B, C
⑤ A, B, C

15. A polynomial $f(x)$ satisfies the given conditions:
$f'(x) = 6x^2 - 4x + 1$ and $f(0) = 3$
What is the value of $f(1)$?
① 1
② 2
③ 3
④ 4
⑤ 5

16. For a function $f(x) = x^2$, the sequence of points which divide the closed interval $[0, 1]$ (including the endpoints) into $n$ segments are $0 = x_0, x_1, x_2, \cdots, x_{n-1}, x_n = 1$ in ascending order. Evaluate $\lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{n}\left\{\frac{f(x_{k-1})+f(x_k)}{2}\right\}$? (Note that $n$ is positive integer equal to or greater than 2.)
① $\frac{1}{3}$
② $\frac{1}{2}$
③ $\frac{2}{3}$
④ $\frac{5}{6}$
⑤ 1

17. Evaluate the definite integral $\int_0^4 \frac{x^2}{x+1} dx - \int_0^4 \frac{1}{x+1} dx$.
① 1
② 2
③ 3
④ 4
⑤ 5

18. Suppose that $\lim_{x\to 1} \frac{1}{x-1} \int_1^x (t^2 + at + 3) dt = 11$. What is the value of a constant $a$?

① 6
② 7
③ 8
④ 9
⑤ 10

19. For all real numbers $x$, a polynomial $f(x)$ is $f(x) + f(-x) = 0$ and satisfies the following conditions. What is the value of $\int_{-3}^{7} f(x)dx$?

| |
|---|
| A. $\int_{0}^{3} f(x)dx = -2$ |
| B. $\int_{-2}^{7} f(x)dx = -5$ |
| C. $\int_{-2}^{0} f(x)dx = 1$ |

① -7
② -4
③ -1
④ 2
⑤ 5

20. What is the area between a curve $y = x^2 - 1$ and a straight line $y = x - 1$?

① $\frac{1}{6}$
② $\frac{1}{3}$
③ $\frac{1}{2}$
④ $\frac{2}{3}$
⑤ $\frac{5}{6}$

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