Florida State University Libraries

2021

Psychometric Report for the 2019 Knowledge for Teaching Early Elementary Mathematics (K-TEEM) Test

Robert C. Schoen, Xiaotong Yang and Gizem Solmaz

Psychometric Report for the 2019 Knowledge for Teaching Early Elementary Mathematics (K-TEEM) Test

Robert C. Schoen
Xiaotong Yang
Gizem Solmaz

APRIL 2021

Research Report No. 2021-04

SECURE VERSION

The research and development reported here was supported by the United States Department of Education through Award Number U423A180115 to Florida State University. The opinions expressed are those of the authors and do not represent views of U.S. Department of Education.

Suggested citation: Schoen, R.C., Yang, X., & Solmaz, G. (2021). *Psychometric Report for the 2019 Knowledge for Teaching Early Elementary Mathematics (K-TEEM) Test* (Research Report No. 2021-04). Florida State University. DOI:10.33009/lsi.1620243057

Copyright 2019, Florida State University. All rights reserved. Requests for permission to use this test should be directed to Robert Schoen, rschoen@lsi.fsu.edu, FSU Learning Systems Institute, 4600 University Center C, Tallahassee, FL, 32306

Detailed information about items are not included in this report. This information was removed in order to release the psychometric report and maintain test security. Requests to view the full report should be directed to Robert Schoen (rschoen@lsi.fsu.edu)

Psychometric Report for the 2019 Knowledge for Teaching Early Elementary Mathematics (K-TEEM) Test

Research Report No. 2021-04

Robert C. Schoen

Xiaotong Yang

Gizem Solmaz

April 2021

Florida Center for Research in Science, Technology, Engineering, and Mathematics (FCR-STEM)

Learning Systems Institute

Florida State University

Tallahassee, FL 32306

(850) 644-2570

Acknowledgements

The research and development reported here was supported by the United States Department of Education through grant award number U423A180115 to Florida State University (Title: *Foundations for Success: Developing Effective Mathematics Educators through Cognitively Guided Instruction*; Principal Investigator: Robert C. Schoen). The opinions expressed are those of the authors and do not represent views of the United States Department of Education.

Robert Schoen designed the content and format of the test, directed the study, created the scoring criteria, interpreted the results, and coordinated the writing of this report. Xiaotong Yang performed the data analysis for the missing-data analysis, dimensionality analysis, item calibration, and item-response theory-based models. Gizem Solmaz assisted with item-response adjudication and the writing and editing of the psychometric report.

Amanda Tazaz created the test form in the Qualtrics platform, provided technical support for participants, reviewed items, and assisted with recruitment of and communication with examinees. She also played an integral role in obtaining permission from the Institutional Review Board at Florida State University to conduct the study and assisting with participant recruitment.

Wendy Bray, Naomi Iuhasz-Velez, Alexandra Utecht, Guillermo Farfan, and Sebnem Atabas scored the teacher response and assisted with item-response adjudication.

Kristy Farina provided support for data management, coordinated assignments for scoring and adjudication of constructed-response items, verified the accuracy of the data, and assisted with description of the scoring criteria.

Insu Paek provided advice throughout the data analysis and interpretation process.

Anne Thistle and Charity Buntin provided valuable assistance with copy editing and formatting.

We are especially grateful for the support from the United States Department of Education and for the hundreds of Florida educators who participated in the study.

Table of Contents

Acknowledgements.....	iv
Executive Summary.....	x
Content and Structure	x
Description of the Sample.....	x
Data Analysis and Scoring	xi
Conclusions and Future Directions	xi
1. Introduction	1
1.1. Item Development	2
1.2. Features of Testing Platform.....	2
1.3 Description of the Study Participants	2
2. Initial Item Review	5
3. Data and Scoring.....	7
3.1 Data Sources	7
3.2 Missing Response Data	7
3.3. Data Entry and Verification Procedures.....	7
3.4. Item Scoring	8
3.5. Item Removal.....	8
4. Dimensionality Analysis	10
5. Classical Test Theory (CTT) Analyses.....	11
5.1. Distribution of the Observed Test Score.....	11
5.2. Item Difficulty and Discrimination	11
5.3. Coefficient α and Standard Error of Measurement	13
6. Item Response Theory (IRT) Analyses.....	14
6.1. Model Description.....	14
6.2. Item Difficulty and Discrimination	14
6.3. Marginal Reliability, Test Information, and Estimated Person Ability	16
6.4. Constructing a Common Scale Between Tests.....	19
References	22

List of Appendixes

Appendix A. 2019 and 2020 K-TEEM Items in Test-Form Order with Scoring Key	23
Appendix B. Administration Instructions at the Start of the Online Assessment.....	58
Appendix C. Further Specification and Exemplars of Scoring Criteria for Constructed Response Items CCK.SMW.6 and SCK.ISS.2.....	59
Appendix D. Proportion of Teacher Responses by Item	66

List of Tables

Table 1.1. Test Blueprint for the 2019 and 2020 K-TEEM Test Form and the Final Scale.....	1
Table 1.2. Self-Reported Characteristics of Teachers (n = 645) Participating in the 2019 K-TEEM Field Test	4
Table 2.1. Test Blueprint for the 2019 K-TEEM Test, Split by Phase in Data Analysis	6
Table 3.1. Missing Response Frequency in the Sample	7
Table 3.2. Item Indexing and Scoring for both Test-Form and Final-Scale Formats.....	9
Table 4.1. Eigenvalues Estimated from Mplus and Their Corresponding Percentages of Explained Variation.....	10
Table 5.1. Item Difficulty and Discrimination from CTT Analyses.....	12
Table 5.2. Distribution of Item Difficulty and Discrimination Estimates for the Items in the Final Scale...	13
Table 6.1. Descriptive Statistics of Discrimination Estimates and Difficulty Estimates of Each Item	14
Table 6.2. Parameter Estimates and Standard Errors for Final-Scale Items Modeled Using 2PL	15
Table D.1. Proportion of Teacher (N=645) Responses by Item	66

List of Figures

Figure 5.1. Bar graph depicting the distribution of the observed test score in the final-scale format.	11
Figure 6.1. Item-difficulty estimate (b) of each final-scale item, presented in sequential order.	16
Figure 6.2. Test information curve and CSEM for the final scale format.	17
Figure 6.3. Person abilities (θ) estimated by maximum likelihood estimation (MLE).	18
Figure 6.4. Person abilities (θ) estimated by expected a posteriori (EAP).	18
Figure 6.5. Person abilities (θ) estimated by maximum likelihood estimation (MLE).	20
Figure 6.6. Person abilities (θ) estimated by expected a posteriori (EAP).	20
Figure 6.7. Person abilities (θ) estimated by maximum likelihood estimation (MLE).	21
Figure 6.8. Person abilities (θ) estimated by expected a posteriori (EAP).	21

List of Equations

Equation 1. Standard Error of Measurement (SEM) from CTT Analyses (1)	13
Equation 2. Two-Parameter (2PL) Model (2)	14
Equation 3. Conditional Standard Error of Measurement (CSEM) Given Person Ability (3)	16

Executive Summary

The *2019 Knowledge for Teaching Early Elementary Mathematics* (2019 K-TEEM) test measures teachers' mathematical knowledge for teaching early elementary mathematics. It is nearly identical to the 2016 K-TEEM (Schoen, Yang, Tazaz, Bray, & Farina, 2019), differing only in that it omitted the items that were not used in scoring of the original 2016 K-TEEM test.

The intended use of the 2019 K-TEEM test is twofold. Its first purpose is to estimate the effect of professional-development programs designed for educators who are responsible for teaching mathematics to students in the early elementary grade levels on those educators' mathematical knowledge for teaching mathematics at those grade levels. The second is to support empirical research into the association between teacher mathematical knowledge for teaching and other facets of the teaching and learning process, including teacher beliefs, instructional practice, and student learning.

The purpose of the present report is to present results of a large-scale field test of the 2019 K-TEEM test. This field test generated response data provided by 649 practicing educators. This report contains information about the test form, a description of the sample, descriptions of the procedures used for data entry, scoring of responses, and analysis of data, recommended scoring procedures, and findings regarding the distribution of test scores and reliability estimates. This report speaks to the substantive and structural elements of validity (Flake, Pek, & Hehman, 2017). Future work will examine the external validity of the test scores, including the ability of the test to detect effects of professional-development programs on teacher knowledge.

Content and Structure

Within the mathematical-knowledge-for-teaching theoretical framework (Ball, Thames, & Phelps, 2008; Schoen, Bray, Wolfe, Tazaz, & Nielsen, 2017), the 2019 K-TEEM contains items designed to measure early elementary teachers' knowledge in five subcategories: common content knowledge, specialized content knowledge, knowledge of content and students, knowledge of content and teaching, and knowledge of content and curriculum. The 2019 K-TEEM test focuses on the topics of number, operations, and algebraic thinking. It is not designed to measure teachers' knowledge or abilities in other aspects of mathematics, such as geometry, measurement, or data analysis. Three items use a constructed-response format; the remaining 28 items that contribute to the final scale use a selected-response format.

Description of the Sample

The 2019 K-TEEM was administered to 645 elementary educators in Florida during 2019. The majority of the examinees ($n = 527$) identified themselves as kindergarten-, first-, or second-grade teachers. Some ($n = 12$) identified themselves as intermediate-grades teachers. The remainder identified themselves as teachers of multiple grade levels. (e.g., kindergarten through second grade, mathematics coaches).

Approximately 10% of the teachers in the sample had attended at least one year of a professional-development program based on Cognitively Guided Instruction (CGI) before the administration of the 2019 K-TEEM, leaving 575 examinees (89%) who had not yet participated in any part of the CGI program.

Data Analysis and Scoring

The 2019 K-TEEM is composed of 31 items contributing to the final-scale score. We analyzed the data using both classical-test-theory and item-response-theory approaches. According to the first, item-difficulty estimates ranged from .17 to .81, and the item-discrimination estimates ranged from .16 to .41. The observed test score ranged from 2 to 31 with a mean of 16.56 and a standard deviation of 5.39; 4 (0.62%) of examinees in this sample received a perfect score of 31. Coefficient α and conditional standard error of measurement were found to be .79 and 2.47, respectively, with the 2019 field-test sample.

For the item-response-theory analyses, we used a two-parameter model for calibration. For the 645-person field-test sample, the person-ability estimates according to the maximum likelihood estimation method ranged from -4.18 to 5 with a mean of 0.03 and a standard deviation of 1.20. Those according to the expected *a posteriori* method ranged from -2.55 to 2.74 with a mean of 0 and a standard deviation of 0.90.

Conclusions and Future Directions

The overall difficulty of the 2019 K-TEEM test and the abilities of the educators in the 2019 field-test sample aligned reasonably well, and the reliability estimates appeared to be sufficiently high for the intended use of the test. The results described in the present report provide an argument in favor of the substantive and structural aspects of validity (Flake, Pek, & Hehman, 2017). Future validation efforts will determine whether the effects of teacher professional-development interventions can be detected in the 2019 K-TEEM scores and whether the 2019 K-TEEM scores are associated with other factors, such as student learning in mathematics.

1. Introduction

The *2019 Knowledge for Teaching Early Elementary Mathematics* (2019 K-TEEM) test measures teachers' mathematical knowledge for teaching early elementary mathematics (Ball, Thames, & Phelps, 2008; Schoen, Bray, Wolfe, Tazaz, & Nielsen, 2017). Within the mathematical-knowledge-for-teaching theoretical framework, the 2019 K-TEEM contains items designed to measure teachers' common content knowledge, specialized content knowledge, knowledge of content and students, knowledge of content and teaching and knowledge of content and curriculum. The 2019 K-TEEM focuses on the domains of number, operations, and algebraic thinking. It is not designed to measuring teachers' knowledge or abilities in other aspects of mathematics, such as geometry, measurement, or data analysis.

The 2019 K-TEEM is very similar to the 2016 K-TEEM, differing from it only in that items that were on the 2016 form but were removed during subsequent data analysis were dropped from the 2019 form. Table 1.1 shows the test blueprint for the 2019 K-TEEM test form and the final scale after data recoding. The blueprint shows the categories measured by the test as well as the number of items corresponding to each category. The 2019 K-TEEM was used to measure teachers' mathematical knowledge for teaching at the beginning and end of the first year of a multiyear, multisite, randomized controlled trial of a teacher professional-development program called Cognitively Guided Instruction (CGI).

Table 1.1. Test Blueprint for the 2019 and 2020 K-TEEM Test Form and the Final Scale

Category and subcategory	Abbreviation	Number of items	
		Test form	Final scale
Common Content Knowledge	CCK		
Meaning of the Equal Sign and Related Notation	ES	4	4
Meaning of Terms Expression/Equation	EE	2	2
Properties of Operations	PO	3	3
Solving Problems in Many Ways	SMW	1	1
Specialized Content Knowledge	SCK		
Interpreting Student Strategies	ISS	6	5
Connecting Models of Mathematical Ideas	CMMI	3	3
Modeling the Structure of a Word Problem	MSWP	3	3
Knowledge of Content and Curriculum	SCK		
Naming Word-Problem Types	NPT	4	4
Knowledge of Content and Students	KCS		
Relative Problem Difficulty	RPD	3	3
Knowledge of Content and Teaching	KCT		
Selecting Word Problems in Service of Specific Instructional Goals	LG	3	3
Total		32	31

Note. One item (SCK.ISS.1) was removed during the data analysis phase.

1.1. Item Development

The items on the 2019 K-TEEM were developed for the 2016 test. See Schoen et al. (2017) and Schoen et al. (2019) for more information about the development process.

1.2. Features of Testing Platform

Like the previous K-TEEM tests, the 2019 K-TEEM test was administered in a Web-based format by means of the Qualtrics (2005–2014) platform. This platform affords a multimedia approach, supporting the use of images and videos in the items. Examinees accessed the test form through a personalized link that was sent by e-mail directly to each individual person. Examinees were validated against a testing database before gaining access to the platform. Information about how to seek technical support was displayed at the bottom of every page and was available throughout the testing window. All questions were formatted on the platform to allow the test to be completed on a computer or a mobile device. All items in the test were displayed one item per page, and a progress bar appeared at the bottom center of the screen.

All questions used forced response; an answer had to be recorded before the examinee was allowed to advance to the next question. After each response was submitted—an active and deliberate action taken by the examinee—the software did not allow the examinee to return to view any previous questions or change any response. If an examinee was unable to complete the assessment in one sitting, the entry validation allowed for ending that session and returning at a later time to continue the test, starting with the first item in the sequence that had not yet been submitted.

1.3 Description of the Study Participants

The educators represented ten public school districts and one private school in Florida. Table 1.2 shows self-reported characteristics of educators participating in the 2019 K-TEEM field test. The average number of years of teaching experience was 12.32 (SD = 8.92). The minimum number of years of teaching experience reported was 0, and the maximum was 42. Almost all (94%) of the participants in the sample identified themselves as female. Most of the educators reported that they taught kindergarten (26%), first grade (31%), or second grade (25%). Intermediate-grades teachers made up a very small fraction of those who responded. Teachers who instructed multiple grade levels composed 14% of the sample.

Many of the teachers had multiple teaching certifications. Elementary Education (78%) was the most common, followed by Prekindergarten/Primary Education at 22%, English for Speakers of Other Languages at 18%, and Exceptional Student Education Certification at 15%.

Most teachers (79%) responded that they had four or more years of teaching experience; only 18% reported three or fewer years, and 2% did not respond. Almost 94% of the teachers had earned a bachelor's (61%) or master's degree (33%).

The sample consisted mostly of classroom teachers (97%); the remaining participants did not respond to the question. The vast majority of the sample worked in regular public schools (94%), whereas (2%) worked in charter schools, and <1% worked in private schools.

Of the teachers in the 2019 sample, 64 (approximately 10%) had attended at least one year of a professional-development program based on Cognitively Guided Instruction (CGI) before completing the

2019 K-TEEM. The CGI program offers up to three years of training. Fewer than 1% of the participating teachers in the 2019 K-TEEM field-test sample had completed two or more years of the CGI program, leaving 575 examinees (89%) who had not participated in any part of the CGI program at the time they completed the test.

Table 1.2. Self-Reported Characteristics of Teachers (n = 645) Participating in the 2019 K-TEEM Field Test

	<i>n</i>	Proportion
Gender		
Female	606	.94
Male	18	.03
Declined to answer	6	.01
Unknown	15	.02
Grade Role		
K	170	.26
1	198	.31
2	159	.25
3	9	.01
5	3	.00
Multiple grade level	91	.14
Unknown	15	.02
Teaching Certification		
Elementary Education	501	.78
Prekindergarten/Primary Education	144	.22
English for Speakers of Other Languages	116	.18
Exceptional Student Education	97	.15
Educational Leadership	40	.06
Preschool Education	20	.03
Middle Grades Mathematics	10	.02
Secondary Mathematics	1	.00
School Principal	2	.00
Others	62	.10
Years of teaching experience		
Four or more	512	.79
Three or fewer	119	.18
Unknown	14	.02
Highest degree earned		
Bachelor's degree	395	.61
Master's degree	216	.33
Professional diploma	15	.02
Professional degree	4	.01
Unknown	15	.02
School Type		
Public	606	.94
Charter	15	.02
Private	2	<.01
Unknown	22	.03
Highest level of participation CGI K–2 program		
CGI K–2 year 1	64	.10
CGI K–2 year 2	6	.01
CGI K–2 year 3	1	<.01
None	575	.89

Note. Proportions may not sum to 1 because of rounding or in the case where multiple answers were accepted.

2. Initial Item Review

The 2019 K-TEEM test contained 32 numbered items, each assigned an item code reflecting the associated subdomain. These items prompted up to 47 responses from the examinees, because three items require multiple responses. Section 3.4 provides more details about item scoring.

The 32 numbered items were coded into 32 dichotomous (i.e., correct, incorrect) variables. The test consisted of 29 selected-response items and 3 constructed-response items. Item SCK.ISS1 was not included in the final scale, for several statistical reasons, which are explained in section 3.4.

To clarify the item-recoding process here, we label the sets of 47, 32, and 31 items the data-entry, test-form, and final-scale formats of the test, respectively. There are 31 items that contribute to the final-scale score, including 28 selected-response items and 3 constructed-response items. Table 2.1 provides a blueprint for the test and includes a map of the correspondence among the data-entry, test-form, and final-scale formats.

Table 2.1. Test Blueprint for the 2019 K-TEEM Test, Split by Phase in Data Analysis

Item	Test-form item #	Final-scale item #
Common Content Knowledge		
CCK.ES.3	9	9
CCK.ES.5	28	27
CCK.ES.7	20	19
CCK.ES.2	23	22
CCK.EE.1	3	3
CCK.EE.2	6	6
CCK.PO.2	27	26
CCK.PO.7	13	12
CCK.PO.9	31	30
CCK.SMW.6	4	4
Specialized Content Knowledge		
SCK.CCMI.3	22	21
SCK.CMMI.2	16	15
SCK.CMMI.4	18	17
SCK.ISS.1	10	
SCK.ISS.2	8	8
SCK.ISS.3	25	24
SCK.ISS.4	14	13
SCK.ISS.5	30	29
SCK.ISS.6	29	28
SCK.MSWP.1	17	16
SCK.MSWP.2	24	23
SCK.MSWP.3	21	20
Knowledge of Content and Curriculum		
KCC.NPT.1	15	14
KCC.NPT.12	7	7
KCC.NPT.14	5	5
KCC.NPT.15	11	10
Knowledge of Content and Students		
KCS.RPD.4	26	25
KCS.RPD.5	32	31
KCS.RPD.6	1	1
Knowledge of Content and Teaching		
KCT.LG.1	2	2
KCT.LG.2	12	11
KCT.LG.5	19	18

3.Data and Scoring

3.1 Data Sources

The web-based 2019 K-TEEM test was completed by 649 educators during spring/summer 2019. The items on the K-TEEM test and scoring key are provided in Appendix A. Administration instructions accompanying the test are provided in Appendix B. Additional specification of the scoring criteria for two constructed-response items are provided in Appendix C.

3.2 Missing Response Data

Four examinees did not provide complete responses. We decided to exclude these four teachers from the data analysis, because they had 75% or more missing responses. Table 3.1 shows the frequency and percentage of missing response(s) in the sample. After the four incomplete responses were excluded, the analytic sample included 645 educators.

Table 3.1. Missing Response Frequency in the Sample

% of Missing response(s)	Frequency	%	Cumulative %
.00	645	99.40	99.40
.75	1*	.15	99.55
.78	1*	.15	99.70
.91	1*	.15	99.85
.94	1*	.15	100.00
<i>Total</i>	649	100.00	

Note.

*teachers excluded from the analysis.

% of Missing response(s) = the percentage of missing response(s) for a given teacher in the sample; frequency = the number of teachers with a given percentage of missing response(s); % = the percentage of teachers who had given percentage of missing response(s); cumulative % = cumulative percentage of teachers who had given percentage of missing response(s).

3.3. Data Entry and Verification Procedures

Teachers accessed test items through a personalized link to an online questionnaire hosted within Qualtrics. Teachers entered their own responses through a combination of text-entry boxes and point-and-click, multiple-choice responses. The response data were exported from Qualtrics to a CSV file, which was then transferred to the SPSS (version 23.0) platform (IBM Corp., 2015) for scoring. Selected-response items were scored by machine within the SPSS platform.

The responses to the constructed-response items were exported to Microsoft Excel and scored by trained members of the scoring committee using the criteria described in Appendix A and further specified in Appendix C. Raters entered their scores into Excel, and those ratings were merged back into the SPSS file. The result was a file with dichotomous (correct/incorrect) variables. This “raw-score” file was then used for subsequent analysis.

After the four responses to item CCK.ES.3 were reviewed by the adjudication committee, the responses to this item were also scored by machine within SPSS, because the review of responses determined that one, and only one, set of four responses in the empirical data was correct.

3.4. Item Scoring

As explained above, the 47 data-entry variables were recoded into 32 test-form variables, because CCK.EE1, CCK.SMW6, and CCK.ES3 required multiple responses. Item CCK.EE1 contained 8 subitems. Item CCK.SMW6 provided 6 fields in which examinees could describe up to 6 ways of solving the problem. Item CCK.ES3 provided 4 blanks for examinees to complete. Appendix C contains the scoring criteria for constructed response items CCK.SMW6 and SCK.ISS2. Item CCK.EE1 is a testlet that includes 8 subitems, and each subitem is a question that prompts a dichotomous (i.e., true, false) response. A teacher who answered all 8 subitems correctly got 1 point, otherwise, 0 point. The 32 test-form variables were coded into 32 dichotomous (i.e., correct, incorrect) variables.

3.5. Item Removal

Item SCK.ISS1 was excluded from the final scale, because this item was negatively correlated with several other items, and it had the lowest factor loading and item-rest correlation. Table 3.2 shows item indexing and scoring of both test-form and final-scale items.

Table 3.2. Item Indexing and Scoring for both Test-Form and Final-Scale Formats

Item-bank code	Test-form item #	Final-scale item #
KCS.RPD6	1	1
KCT.LG1	2	2
CCK.EE1	3	3
CCK.SMW6	4	4
SCK.NPT14	5	5
CCK.EE2	6	6
SCK.NPT12	7	7
SCK.ISS2	8	8
CCK.ES3	9	9
SCK.ISS1	10	
SCK.NPT15	11	10
KCT.LG2	12	11
CCK.PO7	13	12
SCK.ISS4	14	13
SCK.NPT1	15	14
SCK.CMMI2	16	15
SCK.MSWP1	17	16
SCK.CMMI4	18	17
KCT.LG5	19	18
CCK.ES7	20	19
SCK.MSWP3	21	20
SCK.CCMI3	22	21
CCK.ES2	23	22
SCK.MSWP2	24	23
SCK.ISS3	25	24
KCS.RPD4	26	25
CCK.PO2	27	26
CCK.ES5	28	27
SCK.ISS6	29	28
SCK.ISS5	30	29
CCK.PO9	31	30
KCS.RPD5	32	31

Note. Test-form Item # = the item index from the original test; Final-scale item # = the newly generated item number after item SCK.ISS1 was excluded. All items were scored dichotomously: 0, 1.

4. Dimensionality Analysis

To explore the dimensionality of the test, we ran exploratory factor analysis models using

Mplus 8.0 (Muthén & Muthén, 1998–2017) using the weighted least square estimation with mean and variance adjusted method. Table 4.1 shows the eigenvalues estimated by Mplus with the corresponding percentages of explained variation. The largest percentage of variation, 21.95%, is explained by the first factor.

Table 4.1. Eigenvalues Estimated from Mplus and Their Corresponding Percentages of Explained Variation

	Eigenvalue	% variation explained
1	6.80	21.95
2	1.72	5.55
3	1.62	5.21
4	1.56	5.02
5	1.39	4.49
6	1.30	4.19
7	1.20	3.86
8	1.12	3.62
9	1.03	3.32
10	1.01	3.25
11	0.96	3.08
12	0.91	2.92
13	0.88	2.82
14	0.85	2.74
15	0.80	2.58
16	0.75	2.43
17	0.70	2.27
18	0.69	2.21
19	0.65	2.08
20	0.63	2.02
21	0.59	1.90
22	0.55	1.77
23	0.51	1.64
24	0.47	1.52
25	0.45	1.44
26	0.42	1.36
27	0.38	1.22
28	0.34	1.10
29	0.30	0.98
30	0.27	0.86
31	0.19	0.61

Note. Eigenvalue = the eigenvalue associated with a given factor estimated by Mplus; % Variation Explained = the percentage of variation explained by a given factor in the data.

5. Classical Test Theory (CTT) Analyses

5.1. Distribution of the Observed Test Score

We conducted the classical test theory (CTT) analyses using SPSS 26.0 (IBM corp., 2019). Figure 5.1 shows the bar graph depicting the distribution of the observed test scores. The observed test scores ranged from 2 to 31 with a mean of 16.56 and a standard deviation of 5.39; 4 (0.62%) of examinees in this sample received a perfect score of 31. The sample size is 645.

Figure 5.1. Bar graph depicting the distribution of the observed test score in the final-scale format.

5.2. Item Difficulty and Discrimination

Table 5.1 shows the descriptive statistics, item difficulty, and item discrimination estimates of each item. Because all the items were dichotomously coded, the item-difficulty estimates were calculated as the proportions of correct answers, which were equal to the item means. Item discrimination estimates were calculated as the item-rest correlation coefficients (i.e., corrected item-total correlation coefficients). Table 5.2 shows the distribution of the CTT-based item difficulty estimates and item-rest correlations in the final scale.

Table 5.1. Item Difficulty and Discrimination from CTT Analyses

Item-bank code	Final-scale item #	Mean	St. dev.	Item-rest r
KCS.RPD6	1	0.81	0.39	0.25
KCT.LG1	2	0.73	0.45	0.16
CCK.EE1	3	0.29	0.46	0.20
CCK.SMW6	4	0.21	0.41	0.40
SCK.NPT14	5	0.57	0.50	0.21
CCK.EE2	6	0.58	0.49	0.33
SCK.NPT12	7	0.71	0.45	0.16
SCK.ISS2	8	0.49	0.50	0.35
CCK.ES3	9	0.35	0.48	0.40
SCK.NPT15	10	0.32	0.47	0.16
KCT.LG2	11	0.62	0.49	0.30
CCK.PO7	12	0.60	0.49	0.32
SCK.ISS4	13	0.53	0.50	0.30
SCK.NPT1	14	0.47	0.50	0.36
SCK.CMMI2	15	0.48	0.50	0.29
SCK.MSWP1	16	0.70	0.46	0.33
SCK.CMMI4	17	0.32	0.47	0.30
KCT.LG5	18	0.43	0.50	0.23
CCK.ES7	19	0.17	0.37	0.33
SCK.MSWP3	20	0.67	0.47	0.41
SCK.CCMI3	21	0.61	0.49	0.21
CCK.ES2	22	0.42	0.49	0.20
SCK.MSWP2	23	0.63	0.48	0.40
SCK.ISS3	24	0.35	0.48	0.32
KCS.RPD4	25	0.61	0.49	0.36
CCK.PO2	26	0.71	0.46	0.26
CCK.ES5	27	0.62	0.49	0.34
SCK.ISS6	28	0.44	0.50	0.22
SCK.ISS5	29	0.73	0.44	0.36
CCK.PO9	30	0.72	0.45	0.33
KCS.RPD5	31	0.66	0.47	0.27

Note. Final-scale item # = the newly generated item number after item recoding; Mean = item difficulty; Item-rest r = item-rest correlation coefficient (i.e., corrected item-total correlation coefficient), which is the Pearson correlation between the item score and the test score that excludes the item score. All items were scored dichotomously: 0, 1.

Table 5.2. Distribution of Item Difficulty and Discrimination Estimates for the Items in the Final Scale

Value	Number of items
<i>Item difficulty</i>	
>.90	0
.80–.89	1
.70–.79	6
.60–.69	8
.50–.59	3
.40–.49	6
.30–.39	4
.20–.29	2
.10–.19	1
<.09	0
Mean	.53
Median	.58
Standard Deviation	.17
<i>Item-rest r</i>	
.80–1	0
.60–.79	0
.40–.59	4
.20–.39	24
.00–.20	3
Mean	.29
Median	.30
Standard Deviation	.07

5.3. Coefficient α and Standard Error of Measurement

The coefficient α (Cronbach, 1951) of the test was .79. The scale variance was 29.05. According to Equation 1, standard error of measurement (SEM) was calculated to be 2.47, where σ^2 is the test variance, and ρ_{XX} is the coefficient α of the test.

$$SEM = \sqrt{\sigma^2 \times (1 - \rho_{XX})}, \quad (1)$$

6. Item Response Theory (IRT) Analyses

6.1. Model Description

We used flexMIRT 3.51 (Cai, 2017) to perform the IRT analyses. The final-scale test consists of 31 items, and all the items were coded dichotomously. de Ayala (2009) suggested sample size exceeds 1000 for three-parameter models to be used in the IRT calibration. The sample size is 645, and we used the two parameter logistic (2PL) model for the IRT calibration.

Results of flexMIRT indicated that successful convergence was reached in the computation, and the value of $-2\log\text{likelihood}$ was 23874.87. The formula of the 2PL model is shown in Equation 2 according to the parameterization of de Ayala (2009).

$$P_j(\theta) = \frac{\exp [a_j(\theta - b_j)]}{1 + \exp [a_j(\theta - b_j)]}, \quad (2)$$

where a_j is the discrimination index of item j ($j = 1, 2, \dots, J$), b_j is the difficulty index of item j , P_j is the probability of correct answer, θ is the person ability.

6.2. Item Difficulty and Discrimination

Table 6.1 shows the descriptive statistics of the item discrimination estimates and the item difficulty estimates. Table 6.2 presents the item parameter estimates. Figure 6.1 displays the item-difficulty estimates of each item. The item difficulty estimates ranged from -2.37 to 1.96 with a mean of -0.21 and a standard deviation of 1.13 . The item discrimination estimates ranged from 0.40 to 1.37 with a mean of 0.83 and a standard deviation of 0.27 .

Table 6.1. Descriptive Statistics of Discrimination Estimates and Difficulty Estimates of Each Item

	Mean	St. dev.	Minimum	Maximum
Item discrimination (a)	0.83	0.27	0.40	1.37
Item difficulty (b)	-0.21	1.13	-2.37	1.96

Table 6.2. Parameter Estimates and Standard Errors for Final-Scale Items Modeled Using 2PL

Item-bank code	Final-scale item #	Item discrimination	s.e.	Item difficulty	s.e.
		(a)		(b)	
KCS.RPD6	1	0.78	0.15	-2.08	0.36
KCT.LG1	2	0.43	0.12	-2.37	0.66
CCK.EE1	3	0.51	0.12	1.80	0.42
CCK.SMW6	4	1.34	0.18	1.30	0.15
SCK.NPT14	5	0.53	0.12	-0.56	0.20
CCK.EE2	6	0.89	0.13	-0.44	0.12
SCK.NPT12	7	0.44	0.13	-2.11	0.59
SCK.ISS2	8	0.92	0.13	0.03	0.11
CCK.ES3	9	1.16	0.15	0.68	0.12
SCK.NPT15	10	0.40	0.10	1.96	0.53
KCT.LG2	11	0.84	0.13	-0.69	0.14
CCK.PO7	12	0.86	0.13	-0.58	0.13
SCK.ISS4	13	0.76	0.12	-0.17	0.12
SCK.NPT1	14	1.00	0.13	0.12	0.10
SCK.CMMI2	15	0.73	0.11	0.09	0.13
SCK.MSWP1	16	0.98	0.14	-1.02	0.15
SCK.CMMI4	17	0.82	0.13	1.03	0.18
KCT.LG5	18	0.56	0.11	0.56	0.19
CCK.ES7	19	1.12	0.15	1.77	0.22
SCK.MSWP3	20	1.37	0.18	-0.72	0.10
SCK.CCMI3	21	0.53	0.11	-0.93	0.24
CCK.ES2	22	0.51	0.10	0.71	0.22
SCK.MSWP2	23	1.23	0.15	-0.58	0.10
SCK.ISS3	24	0.84	0.12	0.86	0.15
KCS.RPD4	25	1.00	0.14	-0.56	0.12
CCK.PO2	26	0.75	0.13	-1.33	0.23
CCK.ES5	27	0.97	0.13	-0.59	0.12
SCK.ISS6	28	0.53	0.11	0.45	0.19
SCK.ISS5	29	1.14	0.15	-1.10	0.14
CCK.PO9	30	1.03	0.15	-1.13	0.16
KCS.RPD5	31	0.72	0.13	-1.05	0.21

Note. Final-Scale Item # = the newly generated item number after item recoding; s.e. = standard error.

Figure 6.1. Item-difficulty estimate (b) of each final-scale item, presented in sequential order.

6.3. Marginal Reliability, Test Information, and Estimated Person Ability

As reported in the flexMIRT 3.51 (Cai, 2017) IRT output file, the marginal reliability for response pattern scores (Houts & Cai, 2016) was .81.

Equation 3 is the formula showing the relationship between the test information and the conditional standard error of measurement (CSEM), where I is the test information function for a given person ability, and θ is the person ability. The formula used to calculate the CSEM was in accordance with the recommendations made by de Ayala (2009).

$$CSEM(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (3)$$

Figure 6.2 shows the relationship between the test information curve and CSEM of the test given person-ability estimates. According to Figure 6.3, the person-ability estimates were related to low CSEM (i.e., high accuracy of person-ability estimation) when they ranged between -0.80 and 0 , and they were related to high CSEM (i.e., low accuracy of person-ability estimation) when they were smaller than -1.60 or larger than 1.60 .

Figure 6.2. Test information curve and CSEM for the final scale format.

We used maximum likelihood estimation (MLE) to estimate the latent person ability of each individual. For estimation of person ability by flexMIRT, MLE theta scores for zero score and perfect score teachers were set at -5 and 5 , respectively. Figure 6.3 displays the distribution of person ability estimates according to the MLE method. These estimates ranged from -4.18 to 5 with a mean of 0.03 and a standard deviation of 1.20 .

We also used the expected *a posteriori* (EAP) method to estimate the person ability of each individual. Figure 6.4 displays the distribution of person ability estimates according to the EAP method. These estimates ranged from -2.55 to 2.74 with a mean of 0 and a standard deviation of 0.90 .

Figure 6.3. Person abilities (θ) estimated by maximum likelihood estimation (MLE).

Figure 6.4. Person abilities (θ) estimated by expected a posteriori (EAP).

6.4. Constructing a Common Scale Between Tests

Separate from the sample described thus far in this report, we administered the 2019 K-TEEM to 128 educators in spring 2019 (at the start of the randomized controlled trial) and to a sample of 716 educator in spring/summer 2020. We conducted fixed item parameter calibration (Paek & Cole, 2020) using flexMIRT 3.51 (Cai, 2017) to create a common scale between tests. The common items were used as anchor items, which provided the link between the tests. (Note, all the items were anchor items.) The anchor-item parameter estimates created from the 645-person sample were fixed in the subsequent data calibration, so the person ability estimates for the $n_1=645$, $n_2=128$, and $n_3=716$ waves of data collection were on the same scale.

The test forms are exactly the same across all three administrations. Item SCK.ISS1 was on the test form in all three waves of data collection, but we dropped it from the final-scale for all three waves. For the $n_2=128$ and $n_3=716$ samples, we fixed the item parameters of the 31 common items to be the same as the corresponding item parameter estimates used in the final-scale scoring of the data from the $n_1=645$ sample (see Table 6.2 for the item parameter values), and the population mean and variance were freely estimated.

We used MLE and EAP methods for person-ability estimation. For estimation of person ability using flexMIRT, MLE theta scores for zero score and perfect score teachers were set as -5 and 5 , respectively.

Figures 6.5 and 6.6 display the distribution of person ability estimates using MLE and EAP methods for the 128-person sample. All 128 of these teachers had completed the first year of the three-year CGI program for teachers of kindergarten, first, and second grade. The person-ability estimates using MLE method ranged from -2.38 to 5 with a mean of 0.66 and a standard deviation of 1.43 . The person-ability estimates using EAP method ranged from -1.85 to 3.30 with a mean of 0.59 and a standard deviation of 1.08 .

Figure 6.7 and 6.8 display the distribution of person ability estimates using MLE and EAP methods for the spring/summer 2020 administration of the test ($n_3=716$). Approximately two-thirds of the 716 examinees had completed at least one year of the CGI program for grades K–2 or 3–5 at the time of testing. The person-ability estimates using MLE method ranged from -4.63 to 5 with a mean of 0.33 and a standard deviation of 1.32 . The person ability estimates using EAP method ranged from -2.74 to 3.05 with a mean of 0.29 and a standard deviation of 1.01 .

Figure 6.5. Person abilities (θ) estimated by maximum likelihood estimation (MLE).

Figure 6.6. Person abilities (θ) estimated by expected a posteriori (EAP).

Figure 6.7. Person abilities (θ) estimated by maximum likelihood estimation (MLE).

Figure 6.8. Person abilities (θ) estimated by expected a posteriori (EAP).

References

- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389–407.
- Cai, L. (2017). *flexMIRT version 3.51: Flexible multilevel multidimensional item analysis and test scoring* [computer software]. Vector Psychometric Group.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. The Guilford Press.
- Houts, C. R., & Cai, L. (2016). *flexMIRT user's manual version 3.5: Flexible multilevel multidimensional item analysis and test scoring*. Vector Psychometric Group.
- IBM Corp. (2015). IBM SPSS Statistics for Windows, Version 23.0. IBM Corp.
- IBM Corp. (2019). IBM SPSS Statistics for Windows, Version 26.0. IBM Corp.
- Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, 8(4), 1–9.
- Muthén, L. K. and Muthén, B. O. (2017). *Mplus User's Guide* (8th ed.). Muthén & Muthén
- Qualtrics. (2005). *Qualtrics software* (Version 2019). [Computer software]. Qualtrics.
- Paek, I., & Cole, K. (2020). *Using R for Item Response Theory Model Applications*. Routledge.
- Schoen, R. C., Bray, W., Wolfe, C., Nielsen, L., & Tazaz, A. M. (2017). Developing an assessment instrument to measure early elementary teachers' mathematical knowledge for teaching. *The Elementary School Journal*, 118(1), 55–81. <https://doi.org/10.1086/692912>
- Schoen, R. C., Yang, X., Tazaz, A. M., Bray, W., & Farina, K. (2019). *Development and Initial Field Test of the 2016 K-TEEM (Knowledge for Teaching Early Elementary Mathematics) Test* (Research Report No. 2019-01). Florida State University. <https://doi.org/10.33009/fsu.1581610055>

Appendix A. 2019 and 2020 K-TEEM Items in Test-Form Order with Scoring Key

Test-form item #1

KCS.RPD.6

Test-form item #2

KCT.LG.1

CCK.EE.1(a-h)

© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd

A 10x5 grid of gray squares. The squares are arranged in a pattern that suggests a stylized '100%' or a similar numerical representation. The squares are of varying sizes, with some being larger than others, and they are arranged in a way that creates a sense of depth and perspective.

© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd

Test-form item #4

CCK.SMW.6

[Redacted]

[Redacted]

[Redacted]	[Redacted]
[Redacted]	
[Redacted]	
[Redacted]	
[Redacted]	
[Redacted]	

[Redacted]

[Redacted]	[Redacted]	[Redacted]	[Redacted]
[Redacted]	[Redacted]	[Redacted]	[Redacted]

Test-form item #5

SCK.NPT.14

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #6

CCK.EE.2

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

Test-form item #7

SCK.NPT.12

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #8

SCK.ISS.2

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #9

CCK.ES.3

Test-form item #10

SCK.ISS.1

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #11

KCC.NPT.15

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #12

KCT.LG.2

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #13

CCK.PO.7 (Follow-up to KCT.LG.2)

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

Test-form item #14

SCK.ISS.4

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

Test-form item #15

SCK.NPT.1

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #16

SCK.CMMI.2

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

Test-form item #17

SCK.MSWP.1

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #18

SCK.CMMI.4

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #19

KCT.LG.5

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #20

CCK.ES.7

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

Test-form item #21

SCK.MSWP.3

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #22

SCK.CCMI.3

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

Test-form item #23

CCK.ES.2

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #24

SCK.MSWP.2

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #25

SCK.ISS.3

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #26

KCS.RPD.4

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #27

CCK.PO.2

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

[Redacted]

Test-form item #28

CCK.ES.5

Test-form item #29

SCK.ISS.6

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #30

SCK.ISS.5

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #31

CCK.PO.9

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Test-form item #32

KCS.RPD.5

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Appendix B. Administration Instructions at the Start of the Online Assessment

Knowledge for Teaching Early Elementary Mathematics

INSTRUCTIONS FOR COMPLETING THIS QUESTIONNAIRE:

This questionnaire may take you as much as one hour to complete. You may find it useful to have pencil and scratch paper available as you complete this questionnaire.

This is a questionnaire developed by mathematics educators to measure depth of knowledge for teaching early elementary mathematics. You may notice that these questions are designed to align with the way that a teacher needs to know mathematics in order to teach it.

In completing this questionnaire, you should not spend more than 1 or 2 minutes on any question. Imagine you are responding to real classroom situations, and select the answer that most closely matches what you would do, say, or answer at that moment.

Please answer questions based on your own knowledge. In other words, we request that you do not consult any external references (e.g., books, the Internet, or your colleagues) in order to respond to the questions. We are using the honor system to trust that you will comply with this request.

We recommend that you make every effort to complete the questionnaire in one session. If you are interrupted and must step away, your responses will be saved, and you may use the same survey link to finish at a later time. If you return to the survey from the same computer (and you have not cleared the computer's cookies in the interim), you will be able to resume your session, picking up where you left off; otherwise, the survey system will have you start from the beginning as a new session.

Once you have registered your response to each question and pressed the 'Next' button, you will *not* be able to revisit your responses to questions. In addition, please note that the back button (↶) on your internet browser will *not* bring you to the previous question(s).

We are not using the information gathered by this questionnaire to evaluate any person individually. Your name will not be associated with any reporting of these data, and your responses will not be shared with your principal or your district.

Thank you for your participation. If you have any questions or concerns, please contact:

Robert Schoen
Principal Investigator

rschoen@lsi.fsu.edu

Amanda Tazaz
Project Manager

atazaz@lsi.fsu.edu

Appendix C. Further Specification and Exemplars of Scoring Criteria for Constructed Response Items CCK.SMW.6 and SCK.ISS.2

CCK.SMW.6 (Appears as item #4 on 2019 version of K-TEEM)

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

SCK.ISS.2 (Appears as item #8 on on SEED 16.2 version of K-TEEM)

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

<p>[REDACTED]</p>	<p>[REDACTED]</p>
<p>[REDACTED]</p>	<p>[REDACTED]</p>
<p>[REDACTED]</p>	<p>[REDACTED]</p>
<p>[REDACTED]</p>	<p>[REDACTED]</p>

---	--

Appendix D. Proportion of Teacher Responses by Item

Table D.1. Proportion of Teacher (N=645) Responses by Item

Item	Test-form item #	Final-scale item #	Correct response	Most frequent incorrect responses			
			Response	Response	Response	Response	Response
KCS.RPD.6	1	1	(.81)	(.13)	(.04)	(.02)	
KCT.LG.1	2	2	(.73)	(.24)	(.01)	(.01)	(<.01)
CCK.EE.1a	3a		(.64)	(.36)			
CCK.EE.1b	3b		(.69)	(.31)			
CCK.EE.1c	3c		(.99)	(.01)			
CCK.EE.1d	3d		(.80)	(.21)			
CCK.EE.1e	3e		(.53)	(.47)			
CCK.EE.1f	3f		(.97)	(.03)			
CCK.EE.1g	3g		(.66)	(.34)			
CCK.EE.1h	3h		(.75)	(.25)			
CCK.EE.1		3	(.30)	(.71)			
CCK.SMW.6	4	4	(.21)	(.79)			
SCK.NPT.14	5	5	(.57)	(.36)	(.05)	(.02)	(<.01)
CCK.EE.2	6	6	(.58)	(.20)	(.11)	(.11)	
SCK.NPT.12	7	7	(.71)	(.21)	(.05)	(.03)	
SCK.ISS.2	8	8	(.49)	(.51)			
CCK.ES.3a	9		(.38)	(.59)	(.02)	(<.01)	(<.01)
CCK.ES.3b	9		(.96)	(.02)	(.02)	(<.01)	(<.01)
CCK.ES.3c	9		(.38)	(.59)	(.01)	(.01)	(<.01)
CCK.ES.3d	9		(.38)	(.58)	(.03)	(<.01)	(<.01)
CCK.ES.3	9	9	(.35)	(.65)			
SCK.ISS.1	10		(.26)	(.45)	(.15)	(.14)	
KCC.NPT.15	11	10	(.32)	(.34)	(.22)	(.08)	(.04)
KCT.LG.2	12	11	(.62)	(.25)	(.10)	(.03)	
CCK.PO.7	13	12	(.61)	(.20)	(.11)	(.07)	(.02)
SCK.ISS.4	14	13	(.53)	(.12)	(.17)	(.14)	(.04)
SCK.NPT.1	15	14	(.47)	(.29)	(.08)	(.10)	(.06)
SCK.CMMI.2	16	15	(.48)	(.34)	(.17)	(<.01)	
SCK.MSWP.1	17	16	(.70)	(.20)	(.06)	(.05)	
SCK.CMMI.4	18	17	(.32)	(.52)	(.09)	(.04)	(.03)
KCT.LG.5	19	18	(.43)	(.26)	(.27)	(.04)	
CCK.ES.7	20	19	(.17)	(.43)	(.28)	(.09)	(.04)
SCK.MSWP.3	21	20	(.67)	(.19)	(.09)	(.05)	
SCK.CCMI.3	22	21	(.61)	(.35)	(.02)	(.02)	
CCK.ES.2	23	22	(.42)	(.34)	(.22)	(.03)	
SCK.MSWP.2	24	23	(.63)	(.15)	(.12)	(.10)	
SCK.ISS.3	25	24	(.35)	(.56)	(.06)	(.04)	
KCS.RPD.4	26	25	(.61)	(.21)	(.12)	(.03)	(.03)
CCK.PO.2	27	26	(.71)	(.17)	(.07)	(.03)	(.02)
CCK.ES.5	28	27	(.62)	(.21)	(.13)	(.05)	
SCK.ISS.6	29	28	(.44)	(.22)	(.25)	(.09)	
SCK.ISS.5	30	29	(.73)	(.14)	(.07)	(.06)	
CCK.PO.9	31	30	(.72)	(.16)	(.11)	(.01)	
KCS.RPD.5	32	31	(.66)	(.27)	(.03)	(.02)	(.02)

Note. Proportions may not sum to 1 because of rounding; Test-form item # = the item index from the original test; Final-scale item # = the newly generated item number after item recoding.