Technical Report # 1312

Study of the Reliability of CCSS-Aligned Math Measures (2012 Research Version): Grades 6-8

Daniel Anderson

Julie Alonzo

Gerald Tindal

University of Oregon

Published by

Behavioral Research and Teaching
University of Oregon • 175 Education
5262 University of Oregon • Eugene, OR 97403-5262
Phone: 541-346-3535 • Fax: 541-346-5689
http://brt.uoregon.edu

Note: Funds for this data set used to generate this report come from a federal grant awarded to the UO from the Institute of Education Sciences, U.S. Department of Education: Developing Middle School Mathematics Progress Monitoring Measures (R324A100026 funded from June 2010 - June 2014).

Abstract

In this technical report, we describe the results of a study of mathematics items written to align with the Common Core State Standards (CCSS) in grades 6-8. In each grade, CCSS items were organized into forms, and the reliability of these forms was evaluated along with an experimental form including items aligned with the National Council of Teachers of Mathematics (NCTM) Focal Point Standards. The purpose of the experimental measure was to evaluate how previously existing math items functioned empirically relative to the CCSS items. All included NCTM items were previously rated as linked with the CCSS. Analyses included Rasch modeling to explore the difficulty and functioning of both sets of items, classical reliability statistics (Cronbach's alpha, test-retest, and alternate form reliability) and two sets of Generalizability Theory analyses. Overall, results suggest the CCSS items are more difficult than the NCTM items, and test forms made up exclusively of CCSS items operate at less than ideal reliability levels. The ways in which the results of this study have informed revisions to the easyCBM CCSS Math test forms to enhance the reliability while increasing accessibility are discussed .

CCSS Mathematics Reliability: Grades 6-8

Within a Response to Intervention (RTI) framework, the progress of students performing below expectations is generally evaluated through a set of alternate, equivalently difficulty test forms (Deno et al., 2009). Teachers can then base instructional decisions on students' level of achievement at any specific point in time, as well as on their rate of growth over time. However, the validity of these instructional decisions hinges on the technical characteristics of the test forms. If the test forms are not *truly* equivalently difficult, for example, then inappropriate instructional decisions may result based on data that display spurious trends.

The purpose of this technical report is to report on the reliability *within* and *across* the multiple 2012 research version easyCBM CCSS test forms in grades 6-8. We apply classical test theory and generalizability theory methods to explore the reliability of mathematics items written to assess specific areas of the CCSS in mathematics, grades 6-8, as well as items written to assess the NCTM Focal Point standards in mathematics. We apply three classical test theory methods to evaluate the internal, test-retest, and alternate form reliability of test forms assembled from these different math items. Internal reliability is evaluated through Cronbach's alpha, while test-retest and alternate form reliability are evaluated through correlational analyses. We conduct two generalizability theory (G-Theory) analyses to evaluate the proportion of variance associated with students (as intended), as opposed to test items, test forms, or test occasions.

From a classical test theory perspective, the reliability within test forms was evaluated primarily by Cronbach's alpha, which provides an index of the consistency with which the items within a given test form measure a given construct. Alternate form and test-retest reliability analyses provide an index of the reliability across forms by testing the extent to which the test (in

test-retest reliability) or test forms (in alternate form reliability) consistently measure the same underlying construct.

The first G-Theory analysis uses item responses as the outcome. The second analysis uses the total test score as the outcome. The analysis using the total test score as the outcome was conditioned on the number of items to which the student responded, given that differences in number of items to which students respond drives the total *possible* score. That is, without controlling for the number of items with responses by an individual, the assumed total possible score for all students would be 20 even if the student only responded to 10 items. Conditioning the total score on the number of items to which a student responded enables variance uniquely attributable to students/occasions/test forms to be partitioned without missing data confounding the estimates. Variance components from both analyses were then used to explore how changing the number of levels in any of the facets (items, forms, or occasions) changed the reliability of the measurement process. The variance partitioning is often referred to as a *generalizability study*, while the follow up analysis investigating the impact of altering the levels of the facets is referred to as a *decision study* (Brennan, 2001). From a generalizability theory perspective, within-form reliability was evaluated primarily through the *item*form*occasion* analysis, while across-form reliability was evaluated primarily through the (*forms:persons)*valid*occasions*.

Methods

The easyCBM 2012 research version CCSS math measures in grades 6-8 included 13 total test forms per grade – 3 designated for seasonal benchmark screening, and 10 designated for progress monitoring. Five of these 2012 research version test forms from each grade were included in this study. Additionally, an experimental test form was included, which contained 25 items from easyCBM measures aligned with the National Council of Teachers of Mathematics

(NCTM) focal point standards. This combination of forms and items was selected to provide empirical guidance for development of the final version CCSS math measures published in August of 2013.

Procedures and Participants

Two middle schools located in one district in the Pacific Northwest participated in this study. The first school was relatively small, containing approximately 250 students across grades 6-8, while the second school was relatively large, containing approximately 860 students across grades 6-8. All students from both schools participated in this study.

Six total measures were used in this study. All students were administered two test forms during two testing occasions spaced one week apart, for a total of four tests. That is, students were administered two test forms in succession during the first testing occasion, then administered the same two test forms again one week later. A factorial design was used, as displayed in Table 1. Each cell within the table represents a possible testing condition. Elements above and below the diagonal were included as potential testing conditions, leading to 30 possible testing conditions within each grade. All students were randomly assigned to testing conditions. The top 6 rows of the table display the design, while the remainder of the table displays the sample size within each cell by grade.

The factorial design was used primarily to control for test order effects. For instance, if the first test form a student was administered was significantly more difficult than other test forms, the student could potentially be fatigued by the time he or she began the second test form. However, using the factorial design and randomly assigning students to conditions, any potential test order effects would be randomly distributed across conditions.

All tests were administered in a paper-pencil format. Prior to the study, class rosters were sent to the researchers, who then randomly assigned students to testing conditions and printed test forms for all students. Each test form included a cover page that included the student's name, grade, class period, and the day of testing, as well as the teacher's name. Also included in the title page was a "for office use only" field, which included the testing condition. Figure 1 (p. 45) displays an example cover page with all fields left blank. Following the cover page, both tests were presented for whichever testing condition the student was assigned. Between tests was a full-page stop sign, prompting students to not proceed with the second test until instructed to do so. A footer was added to all test forms with a code that indicated the testing condition as a secondary check for accuracy in packaging study materials.

All tests were mailed to the schools for dissemination to teachers. Each teacher received separate packets for each class period taught. Each class period packet contained test forms for only the students within that class period. Separate packets were included for each day of testing, clearly delineated by the packet label (e.g., Period 3, Day 1; Period 3, Day 2). Additional forms were also provided to teachers *without* a student name, but still pre-assigned a specific test format. These blank forms were provided in case any additional students became a part of the teachers' classroom between the time we obtained the class rosters and teachers administered the tests.

An additional teacher information/instructions packet was provided to all teachers. Within this packet was a two-page document providing teachers with an overview of the study and detailed instructions for participation. Teachers were provided a script from which to read during the administration of all tests. The two-page teacher instructions document is presented in Appendix A (p. 47). Teachers were instructed to provide students with 20 minutes to complete

each test during each testing occasion. The importance of administering students the specific pre-assigned test was emphasized. Finally, a manila envelope was provided to teachers to place all answer sheets in after testing. Answer sheets were then stored in a secure location within the school until the completion of the study, at which point the researchers retrieved them.

Measures

Five 2012 research version easyCBM CCSS math forms in each of grades 6-8 were used in this study. For a complete description of the developmental process used, see Anderson, Irvin, Patarapichayatham, Alonzo, and Tindal (2012). All items were written to align with the common core state standards in mathematics for their respective grade (Anderson, Irvin, Alonzo, & Tindal, 2012) and contained three response options. Features of universal design for assessment (see Thompson, Johnstone, & Thurlow, 2002) were emphasized during the development of all items to help increase accessibility. In total, thirteen test forms were created in each grade with three designated for seasonal benchmarking (fall, winter, and spring) and the remaining ten designated for progress monitoring. All test forms were created to be of equivalent difficulty with a Rasch model. In this study, we used progress monitoring test forms six through ten in each grade. Our selection of these measures was purposeful, and was designed to decrease the chances that the student had previously been administered the test form. Generally, students are progress monitored beginning with the first test form, and administered the tests on a regular basis in accordance with an academic intervention. Given that the study took place during the final week of November and the first week of December, it is unlikely that many students had previously received any of the test forms included in the study. The research version CCSS math measures used in this study were available on the District easyCBM system only in SY 2012-2013, as results of this study led to their revision in the spring/summer of 2013.

In addition to the five 2012 research version easyCBM CCSS math measures, we also included an experimental measure. In a study by Irvin, Park, Alonzo, and Tindal (2012), the easyCBM items originally written to the NCTM focal points were rated for alignment with the common core state standards. The 25 items with the highest rated alignment from within each grade were then compiled into a single form for this study. Thus, the experimental form was not an operational test, but instead represented a compilation of easyCBM NCTM items aligned to the common core state standards. The purpose of this experimental form was to empirically explore how items written to the NCTM standards – but rated as aligned with the common core state standards – functioned relative to items explicitly written to and aligned with the common core state standards.

Preliminary Analyses

The 2012 research version CCSS math items were in their first year of operational use during the time of the study. As such, we expected some items to be functioning at levels less than ideal. As a preliminary analysis, we identified potential items for revision or removal. During the summer of 2013, all CCSS test forms were targeted for revisions to enhance the psychometric properties of the tests. The point-biserial correlation was used as the primary statistic for determining item functioning—a classical test theory indicator of item discrimination. *A priori*, we planned to remove the 5 CCSS items within each test form with the lowest point-biserial correlation from all analyses, as these items would likely not be included in future versions of the tests. The revised 20 item CCSS test forms therefore more accurately reflect future versions of the test.

Prior to exploring the reliability of the 2012 research version easyCBM CCSS math test forms, Rasch analyses were also conducted with the revised 20-item test to explore the

functioning of all items. Specifically, we were interested in the functioning of the NCTM items relative to the CCSS items. Rather than placing all items on a new common scale; however, we placed all items on the NCTM scale. That is, we used a chain equating process where we first freely estimated all NCTM items, then used those estimates as anchors to estimate the CCSS items. Thus, the CCSS items were calibrated *relative to* the NCTM scale. By calibrating the CCSS items on the NCTM scale, CCSS item difficulties could be directly compared to the item difficulties of NCTM items. For example, a CCSS item with a calibrated difficulty of 0.3 could be interpreted as being 0.3 logits more difficult than the average NCTM item. Winsteps software, version 3.6.8, was used for all analyses. By default, Winsteps constrains the mean of item difficulties to 0 while person estimates are freely estimated relative to items calibrations.

Primary Analyses

Five analyses were conducted to explore the reliability of the revised 20 item 2012 research version CCSS test forms: three from a classical test theory framework, and two from a generalizability framework. For the classical test theory analyses, we first computed Cronbach's alpha for each test form on each occasion to examine the internal reliability of the measures. Cronbach's alpha was computed for both the full test as well as the revised 20-item test. We also conducted two correlational analyses to explore the test-retest and alternate form reliability of the measures. For test-retest reliability, correlations were conducted within each test form across testing occasions. For alternate form reliability, correlations were conducted within each occasion across test forms. Test-retest and alternate form analyses were conducted only with the revised 20-item test. For all analyses, we set our *a priori* reliability criteria at .9 or above for excellent reliability, .8 to .89 for good reliability, .7 to .79 for moderate reliability, and below .7 for weak reliability.

We also conducted two generalizability theory analyses to examine variance associated with test items, forms, occasions, and the object of measurement – persons. The first analysis involved conducting individual analyses for each test form. The items from within the specified test form were then crossed with persons and occasions. Limiting the analysis to only a single test form resulted in a fully crossed design. However, we were also interested in the variance associated with test forms. To explore this relation, we first created a raw sum score for each student for each test. Students' total raw score for each form was then treated as nested within persons, and crossed with occasion. Figure 2A (p. 46) presents a heuristic for the fully-crossed analysis, item*person*occasion. Figure 2B presents a heuristic for the nested analysis, (forms:students)*occasions. Note that seven distinct sources of variance were estimable for the fully-crossed analysis, including the variance uniquely associated with each facet, as well as all interactions. However, for the nested analysis, only five sources of variance were estimable, because the variance associated with forms was indistinguishable from a form by person interaction. NCTM forms were excluded from the analysis shown in Figure 2B given that the test form was experimental and not intended to be equivalently difficult to all other CCSS forms.

Missing data for the analysis shown in Figure 2A were handled simply by the algorithm. That is, because items were included as a facet, non-responses were simply not estimated. However, for the model shown in Figure 2B missing data had to be handled more explicitly, given that the amount of missing data determined the total possible test score for a student. To control for missing data, we included the number of items to which the student responded as an additional facet in the analysis. Thus, while conceptually the analysis represented the facets displayed in Figure 2B, as these were the variances of interest, in application we estimated a three-facet model, with test forms nested within people, crossed with occasion and the number of

items to which the student responded, or (*forms:persons)*valid*occasions*, where *valid* represents the number of items to which the student responded.

Results

We discuss our results beginning with the preliminary analyses, which (a) identified five items within each test form for removal prior to our primary analyses and (b) compared how the CCSS items functioned relative to the NCTM items. We then discuss the results of our primary analyses, beginning with the results of our classical test theory analyses, followed by the results of each of our two G-Theory analyses.

Preliminary Analyses

Results from the point-biserial analyses, conducted to determine which 5 items should be excluded from subsequent analyses, are displayed in Tables 2-4 for grades 6-8 respectively (pp. 19-21). The five items identified with the lowest point-biserial correlation are displayed in red font.

Results from our Rasch analyses with the revised 20 item test forms are displayed in Tables 5-10 for grade 6 (pp. 22-27), Tables 11-17 for grade 7 (pp. 28-33), and Tables 18-22 for grade 8 (pp. 34-39). For all tables, the NCTM items are displayed first. Because the NCTM items were freely estimated, the values represent the scale for all CCSS items. That is, the values for the CCSS items represent item difficulties relative to the freely estimated NCTM values. Overall, the results of the Rasch analyses suggested that the CCSS items were, on average, 0.79 logits more difficult than the NCTM items in grade 6, 0.37 logits more difficult at grade 7, and 0.19 logits more difficult at grade 8. These values can be interpreted similarly to standard deviation units. For example, the CCSS items were, on average, approximately .79 standard deviations

more difficult than the NCTM items. Generally, items fit the Rasch model expectations well, with mean square outfit values roughly ranging from 0.8 to 1.2.

Primary Analyses

In this section, we present the results of our primary analyses reporting on the reliability of the 2012 research version CCSS math test forms and the experimental NCTM math form. We first discuss the results of three classical test theory analyses, followed by two generalizability theory analyses.

Classical Test Theory Analyses. Cronbach's alpha internal reliability coefficients are displayed by grade, test form, and testing occasion (Day 1 and Day 2) in Table 23 (p. 40). In all instances, the values are displayed for the full and reduced models, with the full model representing the 25-item test and the reduced model representing the revised 20-item test. Across all grades and test forms, the revised test was consistently more reliable. For grade 6, alpha ranged from .63 to .79 for the reduced model for all CCSS forms. The values for the NCTM forms were higher, at .80 and .85 for Day 1 and 2, respectively. Generally, the test forms functioned more reliably during Day 2 than during Day 1. For grade 7, alpha ranged from .58 to .70 for the reduced model for all CCSS forms. The values for the NCTM forms were again higher, at .74 and .79 for Day 1 and 2, respectively. The reliability did not appear to vary by the testing occasion. For grade 8, alpha ranged from .55 to .72 for the reduced model for all CCSS forms. The values for the NCTM forms were on the higher end, at .71 and .76 for Day 1 and 2, respectively. All test forms functioned more reliably during Day 2 than during Day 1, and in some cases quite dramatically so (e.g., forms 6 and 9).

Test-retest reliability coefficients are displayed in Table 24 (p. 41). For grade 6, the CCSS form coefficients ranged from .61 for test form 10 to .73 for form 9. The NCTM form

coefficient was .77. For grade 7, the CCSS form coefficients ranged from .57 for form 8 to .78 for form 10. The NCTM form coefficient was .52. For grade 8, the CCSS form coefficients ranged from .52 for form 9 to .66 for form 7. The NCTM form coefficient was .65.

Alternate form reliability coefficients are displayed in Tables 25-27 for grades 6-8 respectively (pp. 42-44). In each table, values below the diagonal represent the correlation between forms for Day 1, while values above the diagonal represent the correlation between forms for Day 2. For Grade 6, the CCSS form coefficients ranged from .20 to .78 for Day 1, and .43 to .82 for Day 2. The NCTM form correlated with the CCSS forms from .49 to .81 during Day 1 and from .52 to .66 during Day 2. For Grade 7, the CCSS form coefficients ranged from .51 to .75 for Day 1, and .46 to .85 for Day 2. The NCTM form correlated with the CCSS forms from .40 to .64 during Day 1 and from .23 to .85 during Day 2. For Grade 8, the CCSS form coefficients ranged from .44 to .68 for Day 1, and .31 to .81 for Day 2. The NCTM form correlated with the CCSS forms from .37 to .64 during Day 1 and from .43 to .64 during Day 2.

Generalizability Theory. Complete G- and D-study results are displayed in Appendix B (pp. 50-110). The item*person*occasion analyses are displayed for grade 6 in pages 51-68, grade 7 in pages 69-86, and grade 8 in pages 87-104. The (forms:persons)*valid*occasion analyses are displayed for grade 6 in pages 105-106, grade 7 in pages 107-108, and grade 8 109-110.

Grade 6. For the person*item*occasion analysis in grade 6, approximately 8%-12% of the total variance was associated with students across test forms. The student by item interaction was generally the largest source of variance outside of the residual variance. Across test forms, between 16%-27% of the total variance was associated with a student by item interaction. Under the condition of measurement (two test forms on two occasions) the relative reliability, or G-coefficient, ranged from .73 to .79, while the absolute reliability, or Phi-coefficient, ranged from

.71 to .77. Under different conditions of measurement, the relative reliability was estimated as ranging from .60 (for a 15-item test taken on one testing occasion) to .88 (for a 35-item test taken during 3 testing occasions). The absolute reliability ranged from .59 to .87.

For the (forms:persons)*valid*occasion, approximately 38% of the total variance was associated with persons, while 34% was associated with the number of items to which the student responded (denoted as "Valid"). For relative reliability—which partials out variance associated with the number of items to which students responded—the G-coefficient ranged from .89 for one test form taken during one testing occasion to .96 for three test forms taken during three testing occasions. These values, however, were calculated while extrapolating the variance associated with "Valid" to 20. For absolute reliability—which included variance associated with the number of items to which students responded—the phi-coefficients ranged from .50 to .75. Of course, the primary reason for including the number of items responded to as a facet was specifically to partial out, or control, for its variance.

Grade 7. For the person*item*occasion analysis in grade 7, approximately 5%-9% of the total variance was associated with students across test forms. The student by item interaction was generally the largest source of variance outside of the residual variance. Across test forms, between 15%-20% of the total variance was associated with a student by item interaction. Under the condition of measurement (two test forms on two occasions) the relative reliability, or G-coefficient, ranged from .62 to .77, while the absolute reliability, or Phi-coefficient, ranged from .58 to .75. Under different conditions of measurement, the relative reliability was estimated as ranging from .43 (for a 15-item test taken on one testing occasion) to .88 (for a 35-item test taken during 3 testing occasions). The absolute reliability ranged from .41 to .86.

For the (forms:persons)*valid*occasion, approximately 37% of the total variance was associated with persons, while 38% was associated with the number of items to which the student responded (again denoted as "Valid"). For relative reliability—which partials out variance associated the number of items to which students responded—the G-coefficient ranged from .97 for one test form taken during one testing occasion to .99 for three test forms taken during three testing occasions. These values were again calculated while extrapolating the variance associated with "Valid" out to 20. For absolute reliability—which included variance associated with the number of items to which students responded—the phi-coefficients ranged from .49 to .74.

Grade 8. For the person*item*occasion analysis in grade 7, approximately 6%-10% of the total variance was associated with students across test forms. The student by item interaction was generally the largest source of variance outside of the residual variance. Across test forms, between 15%-21% of the total variance was associated with a student by item interaction. Under the condition of measurement (two test forms on two occasions) the relative reliability, or G-coefficient, ranged from .67 to .77, while the absolute reliability, or Phi-coefficient, ranged from .65 to .76. Under different conditions of measurement, the relative reliability was estimated as ranging from .48 (for a 15-item test taken on one testing occasion) to .88 (for a 35-item test taken during 3 testing occasions). The absolute reliability ranged from .47 to .87.

For the (forms:persons)*valid*occasion, approximately 38% of the total variance was associated with persons, while 28% was associated with the number of items to which the student responded (again denoted as "Valid"). For relative reliability—which partials out variance associated the number of items to which students responded—the G-coefficient ranged from .84 for one test form taken during one testing occasion to .94 for three test forms taken

during three testing occasions. These values were again calculated while extrapolating the variance associated with "Valid" out to 20. For absolute reliability—which included variance associated with the number of items to which students responded—the phi-coefficients ranged from .52 to .76.

Discussion

The results of the preliminary analyses suggested the CCSS items were more difficult than the NCTM items. The increased difficulty was not surprising given that the common core math standards in the middle grades outline similar skills to the NCTM standards, but generally earlier than the NCTM standards (see http://www.achieve.org/files/CCSSandFocalPoints.pdf). The measures were also designed to be more difficult than the NCTM items, which were originally written for students performing significantly below expectations (see Anderson, Lai, Alonzo, & Tindal, 2011). Further, teachers and students were in their first year of transitioning to the common core during the time this study was conducted.

During the summer of 2013, the CCSS math tests were revised to include items from both the previously published NCTM and 2012 research version CCSS tests. The revised tests are intended to have an adequate mix of the easier NCTM items and the more difficult CCSS items to reliably assess the mathematics skills of students across a broad ability range. One of the primary purposes of administering progress-monitoring measures is to monitor the progress of students performing below expectations. Yet, if the items are overly difficult, the students will not be able to access the test and demonstrate their competencies. At the same time, if items are too easy, then the test will exhibit ceiling effects and not work well as a screener. Thus, the results of this study served, in part, as the initial groundwork for the construction of the new tests, which took place during the summer of 2013.

The revision of test forms also appears to be necessary given the results of the primary analyses of this study. That is, across all test forms the reliability was lower than ideal. The *a priori* cutoffs for reliability were .90 or above for excellent reliability, .80 to .89 for good reliability, .70 to .79 for moderate reliability, and below .70 for weak reliability. Applying these criteria relative to Cronbach's alpha, two of the five 2012 research version CCSS math forms in grade 6 displayed moderate reliability while the remaining three displayed weak reliability. The results were slightly better at grade 7, where three of the five forms displayed moderate reliability and the remaining two displayed weak reliability. However, at grade 8, all forms displayed weak reliability (including the experimental NCTM form). The alternate form reliabilities ranged from quite poor (r = .20) to quite good (r = .85). Overall, the alternate form reliabilities were quite variable, which likely had to do with the small sample size receiving each combination of test forms (see Table 1).

Exploring the results of the Generalizability theory analyses, it is evident that more items are likely needed for future versions of the test. Not surprisingly, increasing the number of items universally increased the estimated reliability. By including 35 items the tests were often estimated at or close to displaying "good" reliability (i.e., > .80) within a single testing occasion, and were nearly universally estimated as at least "adequate" (i.e., > .70). However, supplementing the test forms with a greater number of highly discriminating items should also increase reliability overall. During the process of revising the easyCBM CCSS middle school math tests, 30 highly discriminating items were selected for inclusion on progress monitoring forms while 45 were selected for inclusion on benchmark screening assessments.

As a final note, it was encouraging to see that the variance associated with the "forms nested in persons" facet was generally quite small. Future studies should explore the unique

variance associated with forms and persons through a fully crossed design to more adequately explore this variance. Regardless, future revisions of the tests will need to be carefully constructed, so as to maintain adequate alternate form comparability while increasing the overall reliability.

References

Anderson, D., Irvin, P. S., Alonzo, J., & Tindal, G. (2012). *The Alignment of the easyCBM Middle School Mathematics CCSS Measures to the Common Core State Standards* (Technical Report No. 1208). Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Anderson, D., Irvin, P. S., Patarapichayatham, C., Alonzo, J., & Tindal, G. (2012). *The Development and Scaling of the easyCBM CCSS Middle School Mathematics Measures* (Technical Report No. 1207). Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Anderson, D., Lai, C. F., Alonzo, J., & Tindal, G. (2011). Examining a grade-evel math CBM designed for persistently low-performing students. *Educational Assessment, 16*, 15-34. doi: 10.1080/10627197.2011.551084

Brennan, R. L. (2001). *Generalizability theory*. New York: Springer.

Deno, S., Reschly, A., Lembke, E., Magnusson, D., Callender, S., Windram, H., et al. (2009). Developing a school-wide progress-monitoring system. *Psychology in the Schools, 46*, 44-55. doi: 10.1002/pits.20353

Irvin, P. S., Park, B. J., Alonzo, J., & Tindal, G. (2012). *The Alignment of the easyCBM Grades 6-8 Math Measures to the Common Core Standards* (Technical Report No. 1230). Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved November 2, 2011, from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html.

Table 1

Factorial Test Administration Design

Test form (f)	6	7	8	9	10	n
6	-	f6f7	f6f8	f6f9	f6f10	f6fn
7	f7f6	-	f7f8	f7f9	f7f10	f7fn
8	f8f6	f8f7	-	f8f9	f8f10	f8fn
9	f9f6	f9f7	f9f8	-	f9f10	f9fn
10	f10f6	f10f7	f10f8	f10f9	-	f10fn
n	f7f6	fnf7	fnf8	fnf9	fnf10	-
Sample size Grade 6						
6	-	7	8	9	0	13
7	15	-	11	17	16	13
8	11	10	-	8	11	12
9	14	15	12	-	8	9
10	16	10	13	11	-	12
n	16	13	12	11	16	-
Sample size Grade 7						
6	-	11	11	19	11	13
7	8	-	17	11	11	14
8	8	12	-	10	11	13
9	12	11	13	-	13	11
10	13	12	13	12	-	14
n	9	10	10	11	4	-
Sample size Grade 8						
6	-	18	18	10	14	10
7	13	-	8	9	9	10
8	17	12	-	14	14	10
9	12	8	8	-	17	13
10	7	12	7	13	-	8
n	10	16	16	12	13	-

Table 2

Grade 6 Test Form Point-Biserial Correlations

Item	Form				
	6	7	8	9	10
1	.277**	.386**	.473**	.269**	.342**
2	.201**	.382**	.283**	.263**	.461**
3	.534**	.358**	.383**	.404**	0.126
4	.617**	.199**	.343**	.366**	.201**
5	.220**	.415**	.198**	.265**	.343**
6	.508**	.431**	.266**	.231**	.301**
7	.480**	.255**	.467**	.395**	.240**
8	.404**	.156*	.319**	.343**	.237**
9	.313**	-0.003	0.137	.268**	0.124
10	.256**	.188**	0.007	.144*	0.081
11	.241**	.416**	.261**	.266**	.442**
12	.530**	.388**	.396**	.487**	.349**
13	.471**	.373**	.404**	0.063	.377**
14	.409**	.335**	.441**	.410**	.323**
15	.248**	.227**	.512**	.407**	.267**
16	.338**	.405**	.253**	.351**	.282**
17	.402**	.385**	.497**	.463**	.445**
18	.346**	.395**	.315**	.424**	.342**
19	.337**	.219**	.520**	.195**	.386**
20	.478**	.252**	.148*	.284**	.409**
21	.322**	.288**	.290**	.510**	.259**
22	0.042	.472**	.314**	.250**	.420**
23	.400**	.518**	.479**	.174*	.154*
24	.228**	0.138	.156*	.245**	0.048
25	.281**	.258**	.507**	.184*	.154*

Note. Items displayed in red font were removed prior to subsequent analyses.

* $p < .05$

** $p < .01$

Table 3

Grade 7 Test Form Point-Biserial Correlations

Item	Form				
	6	7	8	9	10
1	.341**	0.015	.283**	.235**	.392**
2	.225**	.290**	.304**	.346**	.405**
3	.256**	.359**	0.043	.308**	.345**
4	.183**	.163*	.278**	.181**	.475**
5	.268**	.262**	0.13	.312**	.340**
6	.405**	.265**	.243**	.356**	.213**
7	.297**	.325**	.344**	.351**	0.107
8	.221**	.144*	0.045	.216**	.153*
9	0.046	0.134	.186**	.174**	0.04
10	-0.039	0.091	.250**	0.068	0.135
11	.168*	.380**	.444**	.329**	.421**
12	.444**	.398**	.363**	.260**	.369**
13	.262**	.294**	.367**	.379**	.305**
14	.341**	.396**	.496**	.218**	.250**
15	.154*	.339**	.225**	0.042	.366**
16	.372**	.278**	.417**	.411**	.355**
17	.330**	.324**	.269**	.150*	.392**
18	.438**	.177*	0.11	.275**	.398**
19	.405**	.259**	.365**	.293**	.289**
20	0.09	.165*	0.099	.301**	.321**
21	.339**	.239**	.382**	.153*	.385**
22	.521**	.431**	.292**	.342**	.271**
23	.300**	.514**	0.146	.313**	.399**
24	-0.049	.310**	.396**	.417**	-0.053
25	0.134	.245**	.315**	0.098	0.139

Note. Items displayed in red font were removed prior to subsequent analyses.

* $p < .05$

** $p < .01$

Table 4

Grade 8 Test Form Point-Biserial Correlations

Item	Form				
	6	7	8	9	10
1	.343**	.204**	.297**	.295**	.430**
2	.270**	.372**	.350**	.298**	.307**
3	.392**	.302**	.181**	.362**	.271**
4	.327**	.199**	.231**	.289**	.415**
5	.204**	.333**	.139*	.171*	.260**
6	.279**	.256**	.344**	.323**	.349**
7	0.068	.280**	.345**	.378**	0.13
8	.413**	.323**	.283**	.235**	0.101
9	.350**	0.112	.294**	.241**	.427**
10	.346**	.176*	.402**	.192**	0.117
11	.445**	.423**	.361**	.284**	.394**
12	.396**	.351**	.326**	.411**	.328**
13	.242**	.374**	.370**	0.131	.336**
14	.243**	.336**	.375**	.320**	.156*
15	.422**	0.124	.311**	.441**	.381**
16	.420**	.266**	.416**	.367**	.213**
17	.169*	.318**	.317**	.382**	.511**
18	.436**	.386**	.364**	.277**	.484**
19	0.094	.312**	.262**	.392**	.406**
20	.232**	.245**	.328**	.344**	.393**
21	.189**	.426**	.433**	.265**	.311**
22	.283**	.231**	.335**	.173*	.163*
23	.345**	.435**	.279**	.382**	.375**
24	.343**	.223**	0.137	.450**	.279**
25	.433**	.363**	.402**	.343**	.198**

Note. Items displayed in red font were removed prior to subsequent analyses.

* $p < .05$

** $p < .01$

Table 5

Rasch Analyses: Grade 6 NCTM Test Form

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	-1.71	241	220	0.24	0.9	-0.54	0.53	-1.46	0.03	0.37
2	0.21	239	163	0.16	1.03	0.53	1.03	0.26	0	0.39
3	1.22	237	118	0.15	0.85	-2.77	0.81	-2.21	-0.03	0.57
4	2.13	234	79	0.16	1.08	1.05	1.24	1.76	-0.05	0.44
5	1.87	234	89	0.16	0.98	-0.28	1.11	0.98	-0.05	0.5
6	1.3	239	115	0.15	1.25	4.03	1.43	4.23	-0.04	0.3
7	-1.38	238	211	0.22	0.95	-0.28	0.77	-0.73	0.03	0.35
8	-0.07	237	172	0.16	1.02	0.23	0.99	-0.03	0	0.39
9	0.17	233	160	0.16	0.9	-1.47	0.8	-1.64	-0.01	0.49
10	-0.19	239	178	0.16	0.92	-1.06	0.74	-1.8	0	0.47
11	0.11	231	161	0.16	1.1	1.35	1.18	1.34	-0.01	0.35
12	1.13	236	121	0.15	1.2	3.38	1.41	4.09	-0.03	0.32
13	1.09	235	122	0.15	1.22	3.65	1.28	2.91	-0.03	0.32
14	0.43	238	153	0.15	0.89	-1.77	0.8	-1.96	-0.01	0.51
15	0.03	238	169	0.16	1.1	1.37	1.16	1.15	0	0.34
16	-2.32	236	224	0.31	0.99	0.03	0.46	-1.29	0.03	0.28
17	-1.61	236	214	0.24	0.83	-1.01	0.47	-1.82	0.03	0.41
18	0.21	232	158	0.16	1.01	0.24	1.03	0.27	0	0.41
19	-1.08	234	201	0.2	0.92	-0.6	0.83	-0.6	0.03	0.38
20	-0.09	232	170	0.16	0.82	-2.37	0.67	-2.46	0	0.53
21	1.25	229	113	0.15	1	0.04	1.02	0.2	-0.02	0.47
22	0.7	224	133	0.15	1.06	1.06	1.06	0.63	-0.01	0.41
23	-0.81	229	190	0.19	0.92	-0.72	0.81	-0.83	0.02	0.41
24	-1.3	228	201	0.22	0.77	-1.61	0.51	-1.92	0.03	0.47
25	-1.31	227	200	0.22	0.77	-1.62	0.48	-2.06	0.03	0.47

Table 6

Rasch Analyses: Grade 6 Test Form 6

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	0.19	224	150	0.15	1.05	0.84	1.07	0.68	0.01	0.31
2	0.73	221	124	0.15	1.16	2.92	1.23	2.61	0.01	0.23
3	1.53	225	90	0.15	0.86	-2.41	0.86	-1.76	0.01	0.53
4	0.6	221	131	0.15	0.86	-2.66	0.79	-2.56	0.01	0.53
5	1.41	224	95	0.15	0.91	-1.65	0.9	-1.33	0.01	0.5
6	2.32	222	57	0.17	0.93	-0.76	1.12	0.94	0.01	0.42
7	1.71	219	80	0.15	1.03	0.52	1.1	1.1	0.01	0.37
8	1.43	215	91	0.15	1.15	2.49	1.25	2.96	0.01	0.26
9	-0.35	223	171	0.17	1.12	1.38	1.23	1.4	0.01	0.21
10	0.63	222	130	0.15	0.87	-2.44	0.81	-2.29	0.01	0.52
11	1.26	220	100	0.15	1	0.07	0.98	-0.26	0.01	0.41
12	-0.93	222	188	0.2	0.91	-0.71	0.87	-0.55	0.01	0.36
13	0.18	220	148	0.16	0.97	-0.47	0.9	-0.83	0.01	0.41
14	0.03	215	151	0.16	0.95	-0.72	0.91	-0.72	0.01	0.41
15	1.63	209	80	0.16	1.07	1.12	1.14	1.54	0.01	0.33
16	2	207	64	0.17	0.94	-0.73	0.89	-1.01	0.01	0.47
17	0.54	203	122	0.16	1.01	0.19	1.07	0.78	0.01	0.37
18	0.92	203	106	0.15	1.3	4.97	1.48	5.13	0.01	0.11
19	1.6	196	75	0.16	0.99	-0.08	1	-0.02	0.01	0.43
20	1.6	189	72	0.17	1.15	2.1	1.14	1.46	0.01	0.29

Table 7

Rasch Analyses: Grade 6 Test Form 7

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	0.72	239	134	0.14	0.96	-0.78	0.93	-0.99	0.01	0.45
2	2.06	240	71	0.15	1.08	1.08	1.18	1.57	0.01	0.3
3	1.02	237	118	0.14	0.92	-1.59	0.89	-1.63	0.01	0.48
4	0.47	240	147	0.14	0.86	-2.67	0.8	-2.67	0.01	0.54
5	0.85	239	128	0.14	1.06	1.19	1.05	0.71	0.01	0.35
6	1.36	230	98	0.15	1.16	2.85	1.18	2.38	0.01	0.24
7	-0.5	235	185	0.17	0.85	-1.65	0.75	-1.76	0.01	0.48
8	0.25	236	155	0.15	0.96	-0.61	0.93	-0.74	0.01	0.42
9	1.21	232	107	0.14	0.98	-0.32	0.99	-0.07	0.01	0.42
10	1.38	230	98	0.15	1.12	2.12	1.16	2.02	0.01	0.28
11	-0.2	229	170	0.16	0.95	-0.63	0.89	-0.82	0.01	0.41
12	-0.61	223	180	0.18	0.94	-0.6	0.8	-1.19	0.01	0.4
13	0.71	216	124	0.15	0.99	-0.09	0.97	-0.34	0.01	0.4
14	1.06	217	109	0.15	1.2	3.58	1.28	3.72	0.01	0.2
15	0.89	213	114	0.15	1.13	2.29	1.18	2.38	0.01	0.27
16	0.31	201	130	0.16	1.04	0.68	1.05	0.58	0.01	0.34
17	0.33	195	125	0.16	0.9	-1.59	0.85	-1.56	0.01	0.48
18	1.12	192	92	0.16	0.87	-2.46	0.83	-2.4	0.01	0.53
19	2.12	190	53	0.17	1.22	2.48	1.49	3.48	0.01	0.11
20	1.77	190	66	0.17	1.14	1.93	1.15	1.48	0.01	0.25

Table 8

Rasch Analyses: Grade 6 Test Form 8

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	1	200	101	0.16	0.97	-0.5	0.95	-0.62	0.01	0.43
2	-0.39	203	156	0.18	0.94	-0.66	1.23	1.4	0.01	0.34
3	-0.48	202	158	0.18	0.98	-0.19	0.95	-0.23	0.01	0.34
4	0.93	201	104	0.15	0.9	-1.88	0.87	-1.82	0.01	0.5
5	2.28	198	51	0.18	1.1	1.06	1.09	0.68	0.01	0.32
6	1.92	195	62	0.17	1.24	2.9	1.47	3.76	0.01	0.16
7	2.55	192	41	0.19	1.22	1.93	1.7	3.48	0.01	0.14
8	-0.77	202	166	0.19	0.96	-0.31	0.88	-0.58	0.01	0.34
9	0.19	199	133	0.16	0.96	-0.58	0.96	-0.38	0.01	0.4
10	0.21	196	130	0.16	0.92	-1.17	0.93	-0.68	0.01	0.43
11	1.44	193	79	0.16	0.93	-1.21	0.91	-1.12	0.01	0.47
12	1.1	192	92	0.16	0.89	-2.09	0.88	-1.61	0.01	0.5
13	-0.04	190	134	0.17	1.04	0.51	1.12	0.97	0.01	0.3
14	0.32	186	119	0.17	0.89	-1.7	0.86	-1.39	0.01	0.48
15	0.79	188	102	0.16	1.09	1.52	1.09	1.13	0.01	0.31
16	1.39	178	75	0.17	1.23	3.47	1.33	3.55	0.01	0.18
17	-0.3	174	131	0.19	1.06	0.64	1.01	0.1	0.01	0.29
18	0.57	173	102	0.17	1.08	1.3	1.04	0.42	0.01	0.33
19	0.98	169	86	0.17	0.93	-1.19	0.92	-1	0.01	0.47
20	1	161	81	0.17	0.91	-1.55	0.9	-1.19	0.01	0.49

Table 9

Rasch Analyses: Grade 6 Test Form 9

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	1.81	213	80	0.15	1.05	0.81	1.11	1.25	0.01	0.31
2	0.34	215	146	0.16	1.09	1.38	1.08	0.73	0.01	0.26
3	-0.09	213	161	0.17	0.9	-1.17	0.85	-1.1	0.01	0.42
4	0.98	214	118	0.15	0.95	-1.09	0.93	-1	0.01	0.42
5	0.41	213	142	0.16	1	-0.01	0.97	-0.29	0.01	0.35
6	1.02	214	116	0.15	0.96	-0.85	0.92	-1.04	0.01	0.42
7	1.33	209	99	0.15	1.17	3.27	1.2	2.71	0.01	0.2
8	-0.01	214	159	0.17	1.06	0.79	1.03	0.28	0.01	0.27
9	0.88	209	119	0.15	0.86	-2.92	0.82	-2.43	0.01	0.51
10	1.29	206	100	0.15	0.98	-0.38	0.95	-0.66	0.01	0.4
11	1.15	206	106	0.15	0.93	-1.43	0.92	-1.14	0.01	0.44
12	-0.48	209	170	0.19	0.93	-0.66	0.79	-1.17	0.01	0.38
13	0.77	204	121	0.15	0.86	-2.58	0.81	-2.37	0.01	0.51
14	0.09	203	147	0.17	0.93	-0.88	0.91	-0.64	0.01	0.4
15	2.09	201	65	0.16	1.16	2.21	1.29	2.61	0.01	0.2
16	1.38	194	90	0.16	1.02	0.32	1.03	0.38	0.01	0.36
17	-0.02	198	147	0.17	0.85	-1.8	0.79	-1.55	0.01	0.47
18	1.39	193	89	0.16	1.2	3.56	1.32	3.93	0.01	0.16
19	1.42	191	87	0.16	1.16	2.82	1.18	2.3	0.01	0.22
20	2.23	189	56	0.17	1.16	1.97	1.24	1.92	0.01	0.21

Table 10

Rasch Analyses: Grade 6 Test Form 10

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	-0.16	223	150	0.15	0.95	-0.85	0.96	-0.35	0.01	0.39
2	1.13	223	89	0.15	0.89	-2.11	0.85	-2.21	0.01	0.49
3	0.61	218	112	0.15	1.17	3.41	1.19	2.81	0.01	0.17
4	0.65	217	110	0.15	1.15	2.99	1.17	2.59	0.01	0.19
5	1.01	220	94	0.15	0.98	-0.42	0.98	-0.32	0.01	0.39
6	0.21	220	132	0.15	0.98	-0.43	1.04	0.58	0.01	0.36
7	1.91	222	56	0.16	1.08	0.95	1.13	1.08	0	0.24
8	0.01	224	143	0.15	0.9	-1.94	0.83	-2.17	0.01	0.47
9	-0.27	225	156	0.15	0.95	-0.69	0.86	-1.44	0.01	0.4
10	0.71	215	105	0.15	0.98	-0.32	0.97	-0.39	0.01	0.38
11	1.09	211	85	0.15	1.05	1.01	1.15	2.01	0.01	0.27
12	-0.98	214	173	0.18	1.01	0.1	0.92	-0.46	0.01	0.29
13	-0.07	209	136	0.15	0.89	-1.87	0.85	-1.72	0.01	0.47
14	0.18	201	120	0.15	0.97	-0.5	0.99	-0.07	0.01	0.37
15	0.15	198	119	0.15	0.98	-0.37	0.97	-0.37	0.01	0.37
16	1.29	196	70	0.16	0.98	-0.22	1	0.02	0.01	0.37
17	-0.13	191	126	0.16	1.05	0.74	1.1	1.05	0.01	0.27
18	1.05	185	75	0.16	1.18	2.98	1.2	2.64	0.01	0.14
19	1.08	183	73	0.16	1.23	3.71	1.28	3.49	0.01	0.08
20	1.06	180	73	0.16	1.11	1.8	1.16	2.18	0.01	0.21

Table 11

Rasch Analyses: Grade 7 NCTM Test Form

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	-0.46	201	140	0.17	0.95	-0.7	0.86	-1.1	-0.01	0.42
2	0.26	201	112	0.16	0.9	-1.94	0.83	-2	-0.01	0.5
3	-0.09	202	127	0.16	0.91	-1.52	0.84	-1.58	-0.01	0.48
4	-0.5	199	140	0.17	0.95	-0.63	0.93	-0.49	-0.01	0.41
5	1.48	200	63	0.17	1.29	3.52	1.44	3.39	-0.02	0.13
6	-0.37	201	137	0.16	0.92	-1.2	1	0.02	0	0.44
7	-1.51	202	172	0.21	0.95	-0.32	0.72	-1.21	0.01	0.37
8	-0.31	200	134	0.16	1	0.07	0.92	-0.63	-0.01	0.38
9	0.58	200	98	0.16	0.97	-0.52	1	0.01	-0.01	0.43
10	0.39	200	106	0.16	1.12	2.12	1.14	1.67	-0.01	0.29
11	2.09	199	43	0.19	1.15	1.38	1.76	3.72	-0.02	0.17
12	-1.56	196	168	0.22	0.83	-1.28	0.56	-1.97	0.01	0.47
13	1.31	193	67	0.17	0.98	-0.25	1.04	0.41	-0.02	0.42
14	0.26	192	108	0.16	1.06	1.05	1.06	0.69	-0.01	0.34
15	-0.28	190	127	0.17	0.96	-0.59	0.9	-0.85	-0.01	0.42
16	-1.48	189	161	0.22	1	0.03	1.17	0.72	0	0.26
17	-0.95	186	145	0.19	0.86	-1.41	0.72	-1.63	0	0.47
18	2.36	182	33	0.21	1.14	1.1	1.56	2.41	-0.02	0.19
19	-1.16	179	145	0.2	0.84	-1.48	0.85	-0.66	-0.01	0.45
20	0.82	172	76	0.17	1.07	1.12	1.07	0.8	-0.02	0.35
21	-0.64	167	122	0.19	0.82	-2.12	0.74	-1.7	-0.01	0.52
22	-0.94	167	130	0.2	0.75	-2.59	0.58	-2.44	-0.01	0.57
23	-0.37	163	111	0.18	0.95	-0.62	0.86	-1.01	-0.01	0.43
24	0.57	158	79	0.18	1.16	2.43	1.23	2.23	-0.03	0.27
25	0.52	155	79	0.18	1.19	2.86	1.25	2.46	-0.03	0.24

Table 12

Rasch Analyses: Grade 7 Test Form 6

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	0.26	216	117	0.15	0.9	-2.06	0.86	-2.22	0.01	0.48
2	0.89	215	88	0.15	1.01	0.2	0.97	-0.32	0.01	0.35
3	1.59	214	58	0.16	1.09	1.21	1.15	1.21	0.01	0.21
4	-0.93	211	162	0.17	0.79	-2.38	0.65	-2.87	0.01	0.59
5	0.19	210	116	0.15	1.03	0.66	1.04	0.59	0.01	0.32
6	0.26	211	113	0.15	0.93	-1.37	0.91	-1.32	0.01	0.44
7	0.48	211	103	0.15	1.16	3.23	1.14	2.11	0.01	0.19
8	1.74	206	49	0.17	1.23	2.58	1.54	3.38	0.01	-0.03
9	-0.95	207	159	0.18	1.1	1.06	1.01	0.15	0.01	0.26
10	-0.36	204	136	0.16	0.9	-1.49	0.86	-1.52	0.01	0.47
11	0.01	197	116	0.16	1.02	0.41	1.02	0.33	0.01	0.34
12	-0.73	183	134	0.18	0.88	-1.4	0.77	-1.91	0.01	0.5
13	0.07	172	98	0.17	1.01	0.17	1.03	0.43	0.01	0.35
14	0.33	167	85	0.17	0.96	-0.63	0.93	-0.85	0.01	0.41
15	0.06	156	89	0.17	0.99	-0.09	0.98	-0.19	0.01	0.38
16	1.04	153	55	0.18	1.19	2.73	1.28	2.45	0.01	0.12
17	-0.62	150	106	0.19	0.93	-0.72	0.9	-0.74	0.01	0.43
18	-0.35	144	94	0.19	0.87	-1.7	0.85	-1.42	0.01	0.5
19	1.02	135	49	0.19	1.05	0.75	1.16	1.4	0.01	0.27
20	1.06	129	45	0.2	1.16	2.11	1.28	2.19	0.01	0.14

Table 13

Rasch Analyses: Grade 7 Test Form 7

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	0.52	216	94	0.15	0.95	-1.05	0.91	-1.37	0.01	0.42
2	0.58	213	90	0.15	0.93	-1.42	0.9	-1.5	0.01	0.44
3	1.33	211	58	0.16	0.97	-0.34	1	0.05	0.01	0.36
4	-1.07	216	163	0.17	0.97	-0.27	0.97	-0.18	0.01	0.36
5	0.35	207	96	0.15	1.07	1.37	1.08	1.24	0.01	0.28
6	0.45	209	93	0.15	1.18	3.43	1.24	3.49	0.01	0.16
7	-0.2	211	125	0.15	0.93	-1.33	0.88	-1.62	0.01	0.43
8	-0.54	207	137	0.16	0.91	-1.44	0.87	-1.38	0.01	0.43
9	0.11	204	107	0.15	0.95	-1.11	0.97	-0.48	0.01	0.4
10	0.45	205	92	0.15	0.97	-0.55	0.95	-0.73	0.01	0.38
11	0.94	191	66	0.16	1.06	0.92	1.08	0.9	0.01	0.27
12	0.02	186	100	0.16	0.94	-1.31	0.91	-1.43	0.01	0.42
13	0.69	184	72	0.16	1.11	1.94	1.15	1.9	0.01	0.2
14	0.04	176	94	0.16	1.04	0.77	1.02	0.37	0.01	0.3
15	0.93	170	58	0.17	1.12	1.64	1.18	1.78	0.01	0.19
16	-0.19	152	88	0.18	1.06	1.06	1.09	0.99	0.01	0.28
17	-0.04	146	80	0.18	0.9	-1.71	0.86	-1.8	0.01	0.47
18	0.55	146	61	0.18	0.95	-0.78	0.97	-0.28	0.01	0.4
19	0.84	138	48	0.19	1.03	0.36	1.03	0.35	0.01	0.31
20	0.97	138	44	0.19	1.06	0.79	1.09	0.79	0.01	0.26

Table 14

Rasch Analyses: Grade 7 Test Form 8

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	0.03	217	118	0.15	0.97	-0.61	0.97	-0.43	0	0.37
2	0.15	217	112	0.14	1.01	0.18	1.05	0.85	0	0.32
3	-0.45	216	140	0.15	1.04	0.73	1.12	1.48	0	0.25
4	0.06	214	115	0.15	0.95	-1.2	0.92	-1.35	0	0.41
5	0.69	212	85	0.15	1.16	3.04	1.18	2.64	0	0.14
6	0.46	214	96	0.15	0.99	-0.2	0.99	-0.1	0	0.34
7	0.87	209	75	0.15	1.06	1.07	1.07	0.92	0	0.25
8	-0.32	213	132	0.15	0.89	-2.22	0.85	-2.17	0	0.47
9	-0.51	214	141	0.15	0.93	-1.12	0.88	-1.48	0	0.42
10	-0.54	210	139	0.15	0.95	-0.88	0.94	-0.69	0	0.39
11	1.34	208	56	0.16	0.88	-1.63	0.85	-1.38	0	0.45
12	1.01	202	67	0.16	1.03	0.43	1.04	0.52	0	0.29
13	0.09	199	105	0.15	0.96	-0.88	0.93	-1.25	0	0.39
14	-0.12	189	109	0.16	1.04	0.74	1.07	1.04	0	0.29
15	0.57	181	78	0.16	1.11	2.13	1.15	2.28	0	0.18
16	1.04	172	57	0.17	1.08	1.1	1.12	1.24	0	0.21
17	-0.65	163	112	0.18	0.89	-1.54	0.85	-1.41	0	0.45
18	-0.28	157	96	0.17	1	-0.02	0.99	-0.13	0	0.33
19	0.04	146	78	0.18	1.13	2.33	1.12	1.62	0	0.18
20	1.01	141	46	0.19	1	-0.01	1.1	0.92	0	0.31

Table 15

Rasch Analyses: Grade 7 Test Form 9

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	-0.38	229	155	0.15	0.98	-0.36	1.03	0.41	0.01	0.32
2	0.55	226	108	0.14	0.99	-0.11	0.98	-0.28	0.01	0.34
3	1.17	226	78	0.15	0.95	-0.82	0.93	-0.93	0.01	0.38
4	1.23	225	74	0.15	1.06	1.05	1.1	1.23	0.01	0.21
5	-0.48	226	158	0.15	0.97	-0.5	0.94	-0.63	0.01	0.35
6	-0.44	223	155	0.15	0.91	-1.42	0.84	-1.75	0.01	0.42
7	1.24	221	73	0.15	1.06	0.97	1.16	1.9	0.01	0.2
8	-1.32	225	188	0.19	0.89	-0.91	0.75	-1.6	0.01	0.41
9	-0.09	224	139	0.15	0.97	-0.65	0.99	-0.14	0.01	0.35
10	1.34	219	69	0.15	1.02	0.35	1.25	2.63	0.01	0.24
11	0.8	212	90	0.15	1.09	1.96	1.14	2.18	0.01	0.2
12	0.3	216	116	0.14	0.89	-2.64	0.86	-2.56	0.01	0.46
13	0.29	212	115	0.15	1.14	3.22	1.19	3.16	0.01	0.11
14	0.94	198	78	0.15	1.05	0.9	1.05	0.72	0.01	0.26
15	0.71	195	86	0.15	1.03	0.6	1	0.08	0.01	0.29
16	1.29	186	59	0.17	0.94	-0.91	0.96	-0.42	0.01	0.38
17	-0.09	176	108	0.16	1.09	1.66	1.15	1.86	0.01	0.17
18	0.68	167	74	0.16	1.03	0.68	1.1	1.5	0.01	0.26
19	0.58	166	77	0.16	0.97	-0.71	0.95	-0.81	0.01	0.37
20	1.08	151	54	0.18	1.15	2.14	1.21	2.21	0.01	0.12

Table 16

Rasch Analyses: Grade 7 Test Form 10

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	0.36	212	109	0.15	0.93	-1.47	0.91	-1.37	0.01	0.45
2	-0.1	211	129	0.15	0.88	-2.31	0.85	-1.83	0.01	0.47
3	1.32	213	68	0.16	0.98	-0.19	0.98	-0.17	0.01	0.41
4	0.7	209	93	0.15	0.84	-3.12	0.82	-2.69	0.01	0.53
5	1.52	210	60	0.17	0.94	-0.68	1.03	0.26	0.01	0.42
6	-0.72	211	154	0.16	1.07	0.89	1.12	0.95	0.01	0.24
7	1.47	211	62	0.16	1.21	2.5	1.42	3.4	0.01	0.17
8	0.4	205	105	0.15	0.94	-1.3	0.9	-1.44	0.01	0.44
9	-0.58	207	147	0.16	0.95	-0.73	0.89	-0.9	0.01	0.38
10	0.62	200	93	0.15	1.04	0.76	1.05	0.74	0.01	0.34
11	0.98	195	75	0.16	1.06	0.91	1.06	0.7	0.01	0.32
12	0.3	189	100	0.16	0.98	-0.38	0.98	-0.27	0.01	0.38
13	0.36	185	96	0.16	0.98	-0.41	0.97	-0.37	0.01	0.39
14	0.88	180	73	0.16	0.91	-1.45	0.9	-1.27	0.01	0.47
15	0.98	166	65	0.17	1.09	1.4	1.08	0.97	0.01	0.28
16	1.66	166	43	0.19	1.04	0.41	1.19	1.36	0.01	0.3
17	-0.74	157	116	0.19	0.92	-0.86	0.81	-1.26	0.01	0.41
18	0.31	145	75	0.18	0.96	-0.64	0.93	-0.8	0.01	0.41
19	1.13	141	49	0.19	1.37	4.2	1.45	3.8	0.01	-0.04
20	0.68	135	60	0.19	1.15	2.31	1.14	1.68	0.01	0.2

Table 17

Rasch Analyses: Grade 8 NCTM Test Form

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	-0.03	215	115	0.15	0.97	-0.57	0.94	-0.87	-0.01	0.4
2	-0.53	211	135	0.15	1.07	1.26	1.1	1.11	0	0.25
3	-0.09	210	115	0.15	0.99	-0.15	0.99	-0.19	-0.01	0.36
4	0.48	212	91	0.15	1.06	1.12	1.07	1.05	-0.02	0.3
5	-0.23	212	123	0.15	0.83	-3.79	0.78	-3.11	-0.01	0.54
6	-0.1	212	117	0.15	0.9	-2.24	0.85	-2.17	-0.01	0.47
7	-0.48	211	133	0.15	1.11	2	1.33	3.37	0	0.18
8	-0.35	212	128	0.15	1.07	1.33	1.24	2.71	0	0.24
9	-0.25	211	123	0.15	1.16	3.12	1.2	2.41	0	0.17
10	-0.93	213	153	0.16	0.91	-1.34	0.93	-0.54	0	0.41
11	0.45	212	93	0.15	0.87	-2.53	0.86	-2.12	-0.02	0.51
12	0.23	212	102	0.15	0.86	-2.93	0.83	-2.75	-0.01	0.52
13	-0.06	211	115	0.15	0.85	-3.35	0.79	-3.14	-0.01	0.53
14	0.37	210	96	0.15	0.91	-1.84	0.88	-1.8	-0.02	0.48
15	0.89	206	72	0.16	1.25	3.45	1.31	3.34	-0.03	0.1
16	0.2	208	103	0.15	0.84	-3.59	0.79	-3.42	-0.02	0.55
17	0.62	204	83	0.15	1.07	1.13	1.08	1.12	-0.02	0.3
18	0.42	204	91	0.15	0.91	-1.66	0.9	-1.54	-0.01	0.47
19	-0.26	199	118	0.15	0.91	-1.81	0.85	-2.03	-0.01	0.46
20	0.35	200	92	0.15	1.03	0.57	1.04	0.62	-0.02	0.33
21	0.42	193	85	0.16	0.99	-0.15	0.98	-0.23	-0.01	0.38
22	0.55	190	78	0.16	1.24	3.79	1.24	3.03	-0.01	0.11
23	0.33	188	87	0.16	1.07	1.31	1.08	1.13	-0.02	0.29
24	-0.85	188	132	0.17	0.99	-0.1	1	0.06	0.01	0.31
25	-1.15	188	142	0.18	0.93	-0.86	0.89	-0.7	0.01	0.37

Table 18

Rasch Analyses: Grade 8 Test Form 6

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	-0.18	224	125	0.14	1	0.12	0.99	-0.11	0	0.35
2	-0.36	225	134	0.15	0.97	-0.6	0.96	-0.6	0	0.38
3	0.3	226	103	0.14	1.08	1.72	1.15	2.23	0	0.27
4	-0.87	226	158	0.15	1	0.08	0.97	-0.24	0	0.32
5	-0.09	224	121	0.14	0.96	-0.8	1.03	0.47	0	0.38
6	0.08	225	113	0.14	0.99	-0.29	1	0.04	0	0.38
7	-0.87	226	158	0.15	0.9	-1.53	0.8	-2.05	0	0.44
8	0.83	223	78	0.15	1.05	0.78	1.1	1.15	0	0.31
9	-0.26	225	130	0.14	0.95	-0.95	0.94	-0.88	0	0.4
10	-0.16	223	124	0.14	1.03	0.56	1.03	0.45	0	0.32
11	0.06	224	114	0.14	1.1	2.21	1.13	2.13	0	0.25
12	-0.44	222	137	0.15	0.95	-0.94	0.91	-1.21	0	0.4
13	-0.18	219	124	0.15	0.97	-0.67	0.95	-0.71	0	0.38
14	0.26	219	103	0.15	1.02	0.46	1.02	0.42	0	0.34
15	0.61	220	87	0.15	0.99	-0.16	0.96	-0.56	0	0.39
16	1.27	217	58	0.17	1.12	1.41	1.21	1.76	0	0.23
17	-0.31	214	126	0.15	1.08	1.63	1.14	1.89	0	0.25
18	-0.32	211	125	0.15	0.93	-1.39	0.9	-1.41	0	0.42
19	-0.12	208	115	0.15	0.97	-0.69	0.95	-0.77	0	0.39
20	-0.02	205	108	0.15	0.95	-0.97	0.93	-1.11	0	0.41

Table 19

Rasch Analyses: Grade 8 Test Form 7

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	0.8	204	67	0.16	0.95	-0.8	0.96	-0.37	0.01	0.39
2	0.2	202	92	0.15	1.01	0.35	1	0.06	0.01	0.32
3	1.12	204	55	0.17	0.99	-0.1	1.04	0.42	0.01	0.33
4	0.17	203	94	0.15	1.08	1.9	1.15	2.04	0.01	0.21
5	0.13	204	96	0.15	1.02	0.4	1.02	0.32	0.01	0.31
6	0.85	204	65	0.16	1	-0.01	1.04	0.49	0.01	0.33
7	-0.34	204	117	0.15	0.9	-2.36	0.85	-1.94	0.01	0.44
8	-0.43	202	120	0.15	0.96	-0.82	0.93	-0.86	0.01	0.36
9	0.27	203	89	0.15	0.96	-0.77	0.95	-0.63	0.01	0.37
10	0.28	201	88	0.15	1.01	0.33	0.99	-0.11	0.01	0.32
11	0.45	201	81	0.15	1.1	1.97	1.11	1.48	0.01	0.21
12	0.31	201	87	0.15	1.04	0.77	1.03	0.43	0.01	0.29
13	0.29	200	87	0.15	0.93	-1.46	0.9	-1.39	0.01	0.41
14	0.58	200	75	0.16	1.02	0.38	1	-0.03	0.01	0.32
15	1.07	200	56	0.17	1.07	0.88	1.07	0.63	0.01	0.26
16	-0.39	196	115	0.15	0.93	-1.5	0.89	-1.33	0.01	0.4
17	-0.22	195	107	0.15	1.09	1.92	1.08	1.01	0.01	0.22
18	-0.25	191	106	0.15	0.88	-2.71	0.84	-2.08	0.01	0.46
19	0.83	191	62	0.16	1.08	1.21	1.17	1.74	0.01	0.23
20	0.14	187	88	0.16	0.97	-0.55	0.94	-0.89	0.01	0.37

Table 20

Rasch Analyses: Grade 8 Test Form 8

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	0.56	222	95	0.15	1.06	1.1	1.04	0.56	0.01	0.33
2	-0.21	222	131	0.15	0.95	-1.04	0.91	-1.04	0.01	0.41
3	-0.58	221	148	0.15	0.95	-0.79	0.9	-0.99	0.01	0.38
4	-0.09	220	125	0.15	0.98	-0.52	0.94	-0.73	0.01	0.38
5	1.25	221	65	0.16	1.1	1.3	1.22	1.96	0.01	0.26
6	0.58	221	94	0.15	1	0.08	0.99	-0.17	0.01	0.37
7	0.3	221	107	0.15	1	0.07	1	0.08	0.01	0.36
8	-0.93	221	162	0.16	0.96	-0.49	0.9	-0.74	0.01	0.35
9	-0.16	219	128	0.15	1	0.07	1.02	0.21	0.01	0.34
10	0.61	219	92	0.15	0.97	-0.54	0.94	-0.83	0.01	0.41
11	0.31	217	105	0.15	0.96	-0.78	0.93	-0.95	0.01	0.41
12	0.56	216	93	0.15	1.06	1.13	1.03	0.46	0.01	0.32
13	0.13	213	111	0.15	0.94	-1.32	0.93	-0.93	0.01	0.42
14	0.84	210	79	0.15	1.06	1.03	1.09	1.04	0.01	0.32
15	-0.11	206	118	0.15	1.02	0.5	0.97	-0.3	0.01	0.34
16	0.76	202	79	0.16	1.06	1.01	1.09	1.02	0.01	0.32
17	-0.35	201	126	0.16	0.89	-2	0.95	-0.53	0.01	0.44
18	-0.51	199	131	0.16	0.97	-0.42	1.01	0.09	0.01	0.36
19	-0.34	199	124	0.16	1.06	1.08	1.06	0.65	0.01	0.3
20	0.19	197	100	0.15	0.98	-0.49	0.98	-0.21	0.01	0.39

Table 21

Rasch Analyses: Grade 8 Test Form 9

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	-0.86	202	136	0.16	1	0.1	1.01	0.14	0	0.3
2	-0.69	202	129	0.16	0.99	-0.25	1.2	1.87	0	0.31
3	0.06	203	97	0.15	0.97	-0.58	0.94	-0.88	0	0.4
4	0.27	203	88	0.15	1.07	1.3	1.08	1.09	0	0.29
5	0.11	203	95	0.15	1.04	0.77	1.03	0.44	0	0.32
6	0.31	202	86	0.15	0.99	-0.13	0.97	-0.36	0	0.38
7	0.21	201	90	0.15	1.11	2.01	1.16	2.06	0	0.25
8	0.16	201	92	0.15	1.07	1.31	1.1	1.33	0	0.29
9	0.12	202	94	0.15	0.92	-1.57	0.98	-0.23	0	0.43
10	-0.13	202	105	0.15	1.01	0.32	0.99	-0.05	0	0.34
11	1.18	200	52	0.17	0.95	-0.48	0.92	-0.6	0	0.43
12	0.36	198	82	0.16	0.99	-0.18	0.96	-0.53	0	0.39
13	0.09	199	94	0.15	1.01	0.12	0.99	-0.12	0	0.36
14	0.55	197	74	0.16	1.08	1.22	1.07	0.84	0	0.3
15	0.67	196	69	0.16	0.91	-1.29	0.86	-1.64	0	0.48
16	0.28	194	84	0.16	1.05	0.92	1.04	0.57	0	0.32
17	-0.32	191	107	0.16	1.05	1.07	1.05	0.6	0	0.29
18	-0.35	187	106	0.16	0.97	-0.52	0.92	-0.88	0	0.38
19	0.47	187	73	0.16	0.95	-0.74	0.93	-0.85	0	0.43
20	0.29	185	79	0.16	1.03	0.49	1.07	0.96	0	0.33

Table 22

Rasch Analyses: Grade 8 Test Form 10

Item	Difficulty	Count	Score	Error	Infit MS	Infit MS Z	Outfit MS	Outfit MS Z	Displacement	Pt. Msr.
1	0.78	210	84	0.15	0.94	-1.08	0.91	-1.14	0.01	0.44
2	-0.28	210	131	0.15	1.03	0.64	1.14	1.6	0.01	0.3
3	-0.25	206	127	0.15	1.01	0.14	0.95	-0.57	0.01	0.35
4	0.17	210	111	0.15	1	0.1	1.02	0.35	0.01	0.36
5	0.8	208	82	0.15	1.15	2.5	1.18	2.13	0.01	0.23
6	0.15	209	112	0.15	1.01	0.3	1.02	0.37	0.01	0.35
7	0.79	208	83	0.15	0.98	-0.4	0.98	-0.23	0.01	0.4
8	-0.2	208	127	0.15	0.96	-0.79	0.9	-1.24	0.01	0.4
9	0.64	208	90	0.15	1.01	0.23	1.02	0.27	0.01	0.36
10	0.56	207	93	0.15	1.07	1.39	1.11	1.53	0.01	0.3
11	0.75	200	83	0.16	0.97	-0.52	0.93	-0.88	0.01	0.41
12	0.51	199	92	0.15	1.09	1.66	1.1	1.47	0.01	0.28
13	0.42	196	94	0.15	0.87	-2.76	0.83	-2.6	0.01	0.5
14	0.83	193	76	0.16	0.93	-1.24	0.9	-1.21	0.01	0.45
15	0.6	196	87	0.16	0.97	-0.52	0.96	-0.58	0.01	0.4
16	1.17	189	62	0.17	1.01	0.13	1.02	0.2	0.01	0.37
17	-0.35	187	120	0.16	1.06	0.93	1.03	0.36	0.01	0.29
18	0	186	106	0.16	0.97	-0.59	0.95	-0.61	0.01	0.39
19	0.71	182	76	0.16	1.08	1.42	1.1	1.29	0.01	0.29
20	0.88	181	69	0.17	1.16	2.47	1.34	3.49	0.01	0.21

Table 23

Cronbach's Alpha Reliability Coefficients

		Alpha			
		Day 1		Day 2	
Grade	Form	Full model	Reduced Model	Full model	Reduced Model
6	6	.70	.72	.77	.79
6	7	.66	.69	.67	.72
6	8	.69	.76	.74	.78
6	9	.65	.70	.61	.65
6	10	.57	.63	.59	.69
6	NCTM	.80	-	.85	-
7	6	.57	.68	.63	.69
7	7	.63	.70	.58	.63
7	8	.57	.66	.58	.68
7	9	.58	.58	.51	.59
7	10	.61	.69	.62	.68
7	NCTM	.74	-	.79	-
8	6	.51	.58	.68	.72
8	7	.58	.61	.60	.62
8	8	.61	.65	.70	.71
8	9	.55	.55	.67	.72
8	10	.63	.65	.66	.71
8	NCTM	.71	-	.76	-

Table 24

Test-Retest Reliability Coefficients

Grade	Form	Test-Retest *r*
6	6	.69
6	7	.69
6	8	.71
6	9	.73
6	10	.61
6	NCTM	.77
7	6	.73
7	7	.75
7	8	.57
7	9	.61
7	10	.78
7	NCTM	.52
8	6	.62
8	7	.66
8	8	.65
8	9	.52
8	10	.62
8	NCTM	.65

Table 25

Grade 6: Alternate Form Reliability Coefficients

Test form	6	7	8	9	10	n
6	-	.432	.601	.597	.465	.662
7	.376	-	.819	.641	.760	.572
8	.721	.525	-	.813	.744	.591
9	.492	.720	.426	-	.752	.522
10	.197	.784	.553	.728	-	.549
n	.806	.491	.665	.743	.569	-

Note. Coefficients below the diagonal represent correlations from the first testing occasion, while the coefficients above the diagonal represent correlations from the second testing occasion occurring one week later.

Table 26

Grade 7: Alternate Form Reliability Coefficients

Test form	6	7	8	9	10	n
6	-	.848	.723	.470	.464	.233
7	.542	-	.538	.467	768	.766
8	.746	.514	-	.666	.670	.706
9	.576	.591	.547	-	.845	.602
10	.678	.690	.711	.609	-	.850
n	.599	.566	.400	.642	.612	-

Note. Coefficients below the diagonal represent correlations from the first testing occasion, while the coefficients above the diagonal represent correlations from the second testing occasion occurring one week later.

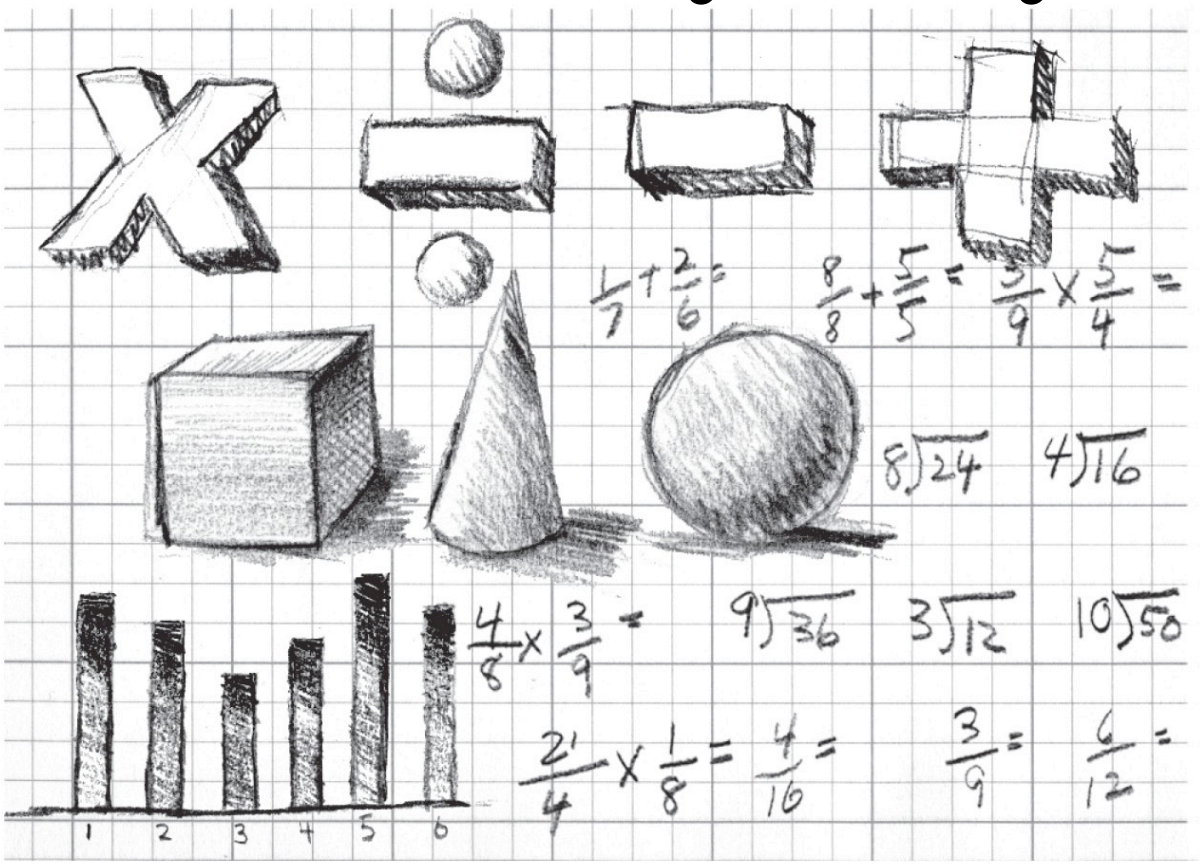
Table 27

Grade 8: Alternate Form Reliability Coefficients

Test form	6	7	8	9	10	n
6	-	.631	.582	.790	.560	.637
7	.560	-	.798	.313	.771	.554
8	.555	.529	-	.809	.597	.558
9	.655	.436	.759	-	.682	.628
10	.566	.675	.575	.528	-	.428
n	.622	.468	.367	.643	.516	-

Note. Coefficients below the diagonal represent correlations from the first testing occasion, while the coefficients above the diagonal represent correlations from the second testing occasion occurring one week later.

CCSS Study: Day 1



Teacher:

Course/Period:

Student:

Grade: For office use only:

*Figure 1.*Example cover page for test forms. All fields were filled in prior to test forms being mailed.

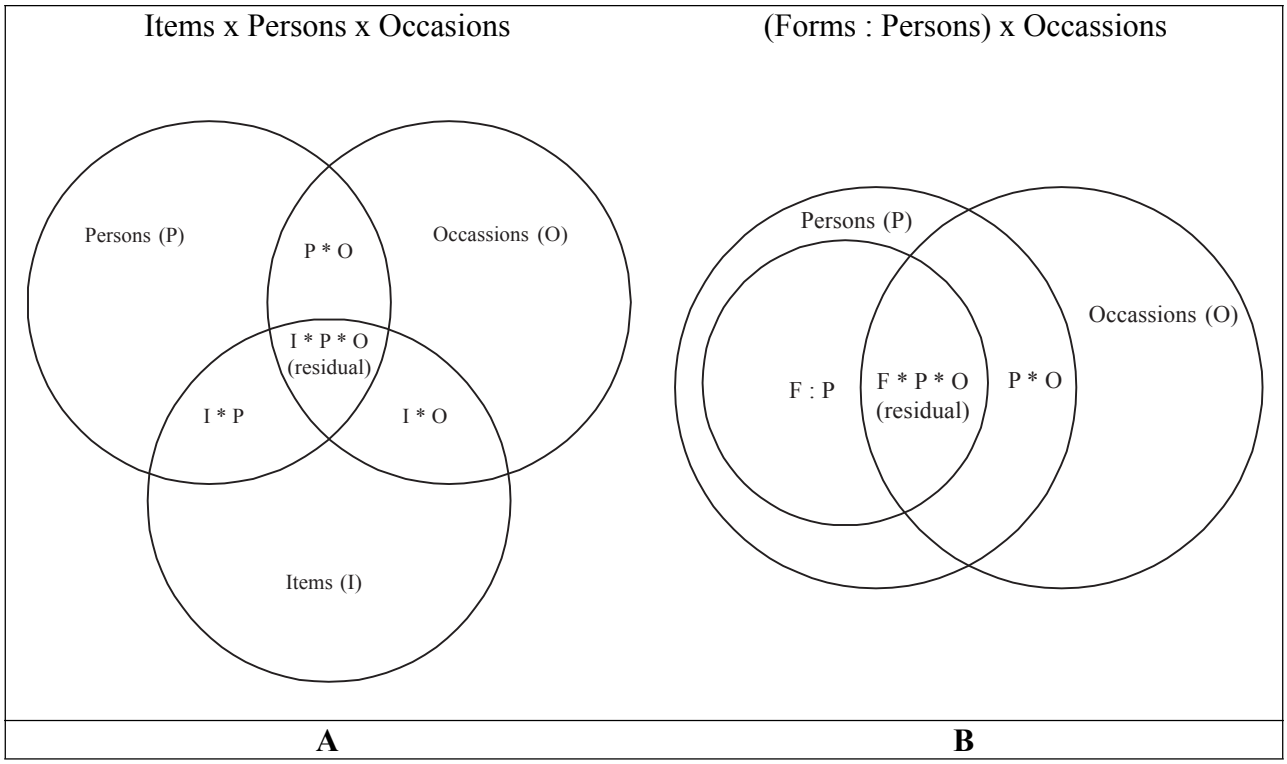


Figure 2. Generalizability theory variance component heuristics. Note that the model in A was conducted for each test form individual, while the model in B was conducted across all test forms.

Appendix A: Teacher Instructions

Teacher Instructions: CCSS Math Reliability Study (Study #4): Grades 6-8

Thank you for agreeing to take part in this study! The purpose of this research is to examine the consistency of students' math test scores regardless of the test form administered, order of administration, or day the testing occurs.

This study will take place over the course of two days, spaced one week apart. During each day we ask that you administer two tests in succession to your students. These two tests will be combined in a single test booklet, with different students in your class receiving different test forms, based on our random assignment of test form combinations across students. Please note: your students will be repeating the same tests on Day 2 of the testing that they took on Day 1 of the testing. This is not a mistake; it is intentional so that we can compare performance on the same test form across two different test times. Below are the directions for participating in this study.

Study outline

Day 1	**Administration Time**
Day 1: Test 1	20 minutes
Day 1: Test 2	20 minutes
Day 2 (one week later)	
Day 2: Test 1	20 minutes
Day 2: Test 2	20 minutes

Step 1: Organize the materials and plan the testing sessions
Materials: You should have received in your packet the following materials:

- 1 copy of the Teacher Instructions in a manila envelope (this document)
- 2 manila envelopes for returning the answer sheets (one for each day)
- 2 groupings of manila envelopes filled with testing materials for each day of testing. Inside each are:
    - One pre-assigned test packet for each student, with his or her name labeled.
    - One pre-assigned scan sheet for each student, with his or her name labeled (paper clipped to each student's test packet).
    - Extra test packets with no student names pre-assigned
    - Extra scan sheets with no student names pre-assigned (paper clipped to unassigned test packets)

Students will need #2 pencils to complete the answer sheets. You should plan for two days of testing, spaced one week apart. Materials from one set of manila envelopes will be used for each testing occasion. The envelopes are marked either Day 1 or Day 2. Please use the correct day's envelopes.

Note. **All students have been pre-assigned DIFFERENT test packets. Please make sure students receive EXACTLY the test packet they were pre-assigned, and that they answer all questions on their pre-assigned answer sheet.**

If there are any students in your class who do not have a pre-assigned test packet, please use one of the blank packets. Place the student's name on both the test packet and the attached (paper clipped) answer sheet. Please **make sure** the student uses the answer sheet attached, so we know which test packet the student was administered when you return all the answer sheets to the university.

Step 2: Administer the tests
Each day of testing will include two tests with 25 questions each, for a total of 50 questions. The testing packet for each student includes both tests, with a page in the middle asking students to "stop".

Blank test forms: In each packet for each of your class periods, there are blank forms with no student name. These are in case you have new students in your class that were not listed on the class rosters we received. Should you need to use these, please **MAKE SURE** that you use the same test form each day. To do this, you look at the "For office use only" number on the cover sheet. For instance, on Day 1 "For office use only" the number may be something like "f6f7". For Day 2, please make sure the student who received the "f6f7" test again receives the "f6f7" test materials. The blank test forms should all be in the same order for Day 1 and Day 2, but please be very careful of this. The integrity of the study rests on students receiving the same tests in the same order on both days.

On day 1, please begin by distributing the test packets from the "Day 1" manila envelopes to all students. Place the materials face down. Ask them to not start until you say. "Begin".

Say the following:

"Today we will be taking two math tests back-to-back. When I say "Begin" you will have 20 minutes to complete Test 1. Please answer all questions using the answer sheet [hold up example scan sheet] **by completely filling in the appropriate bubble with a #2 pencil."**

"When you are finished with Test 1, you will reach a stop sign. Please do not go beyond the stop sign. Once the 20 minutes are up, I will say, 'Stop'. Please put your pencils down at this time, even if you have not yet finished the entire test. We will then begin Test 2. You will also have 20 minutes to complete Test 2."

"Please do your best work. If you are not sure what the right answer is, just choose the answer that you think makes the most sense. Please do not rush to finish each test. It is more important that you do your best than that you finish the test. However, it is also important to not spend too long on any one question."

"While you are working, I can not answer any questions about any of the problems. When you are finished, please turn your paper over, take out your book, and read quietly. Are there any questions? [Answer any questions] **You may begin."**

Start a timer. After 20 minutes have passed, please say, **"Stop"**.

Have all students turn to the beginning of Test 2 on page 9. Once everybody is ready, say, **"You may begin Test 2."**

Start the timer again. After 20 minutes have passed, please say, **"Stop"** and collect all testing materials from students.

Please repeat this process exactly on day 2 of testing, using the material from the Day 2 envelopes.

Step 3: Return the answer sheets

After the first day of testing has completed, please gather the scan sheets (NOT the tests) and place them in the "Day 1 results" manila envelope. Please then take this package to the office and drop them off with the secretary. He or she should have a secure box to hold all scan sheets, which we will be picking up at the end of the study.

Thanks again so much for participating! As compensation, you should receive $150 from your district for the purchase of classroom supplies.

Appendix B:
Complete Generalizability and Decision Study Results

Grade 6 Form 6

Variance Component	Proportion
Student	.096
Day	.000
Item	.039
Student * Item	.163
Student * Day	.006
Day * Item	.001
Student * Day * Item (residual)	.696

Error Variances:

Relative	*Absolute*
.007	.008

G-coefficients:

G	*Phi*
.771	.759

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.016	.013	.010	.009	.008
2	.010	.008	.006	.005	.005
3	.008	.006	.005	.004	.004

D-Study Item Relative Error Variances

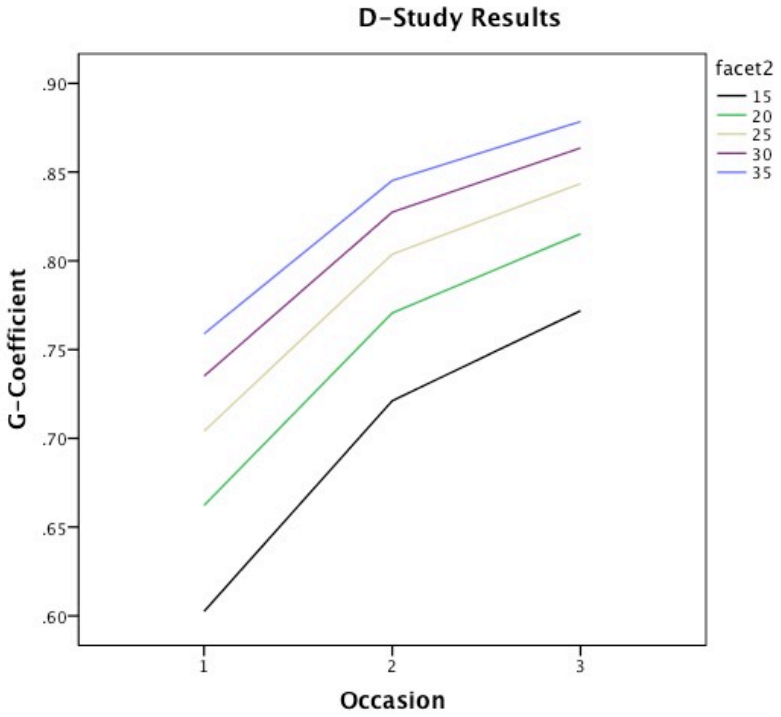
Forms	*n* Items				
	15	20	25	30	35
1	.016	.012	.010	.009	.008
2	.009	.007	.006	.005	.004
3	.007	.005	.004	.004	.003

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.602	.662	.704	.735	.759
2	.721	.771	.804	.827	.845
3	.772	.815	.844	.864	.879

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.592	.653	.696	.728	.752
2	.707	.759	.793	.818	.837
3	.756	.802	.832	.853	.869


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 6 Form 7

Variance Component	Proportion
Student	.102
Day	.000
Item	.069
Student * Item	.250
Student * Day	.001
Day * Item	.002
Student * Day * Item (residual)	.576

Error Variances:

Relative	*Absolute*
.007	.008

G-coefficients:

G	*Phi*
.787	.765

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.015	.012	.009	.008	.007
2	.010	.008	.006	.005	.005
3	.009	.007	.005	.004	.004

D-Study Item Relative Error Variances

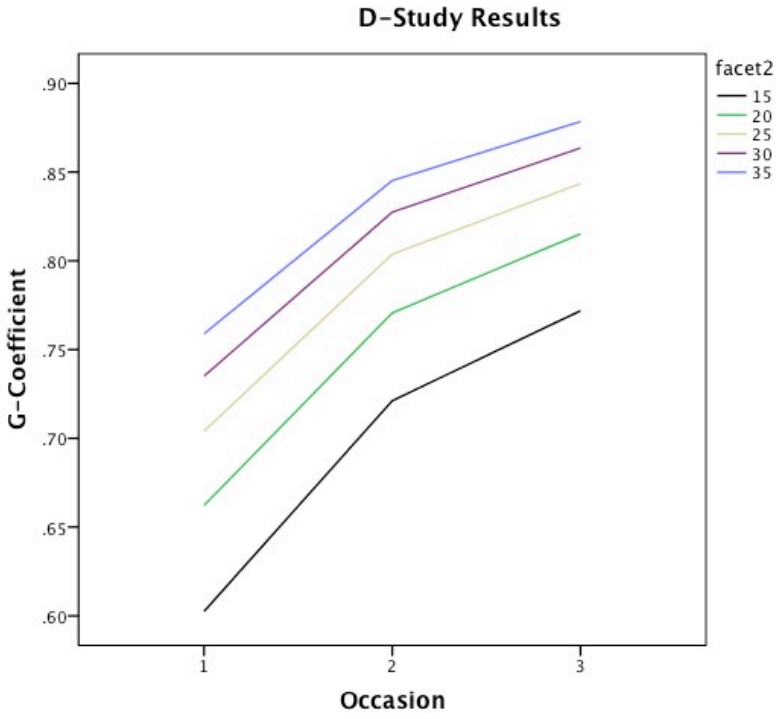
Forms	*n* Items				
	15	20	25	30	35
1	.014	.011	.009	.007	.006
2	.009	.007	.006	.005	.004
3	.007	.006	.005	.004	.003

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.643	.704	.747	.778	.802
2	.736	.787	.821	.845	.863
3	.773	.818	.849	.870	.886

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.623	.685	.729	.762	.787
2	.710	.765	.801	.827	.847
3	.746	.795	.828	.852	.870


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 6 Form 8

Variance Component	Proportion
Student	.116
Day	.000
Item	.074
Student * Item	.273
Student * Day	.012
Day * Item	.001
Student * Day * Item (residual)	.524

Error Variances:

Relative	*Absolute*
.008	.009

G-coefficients:

G	*Phi*
.780	.761

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.017	.013	.011	.010	.009
2	.011	.009	.007	.006	.006
3	.009	.007	.006	.005	.005

D-Study Item Relative Error Variances

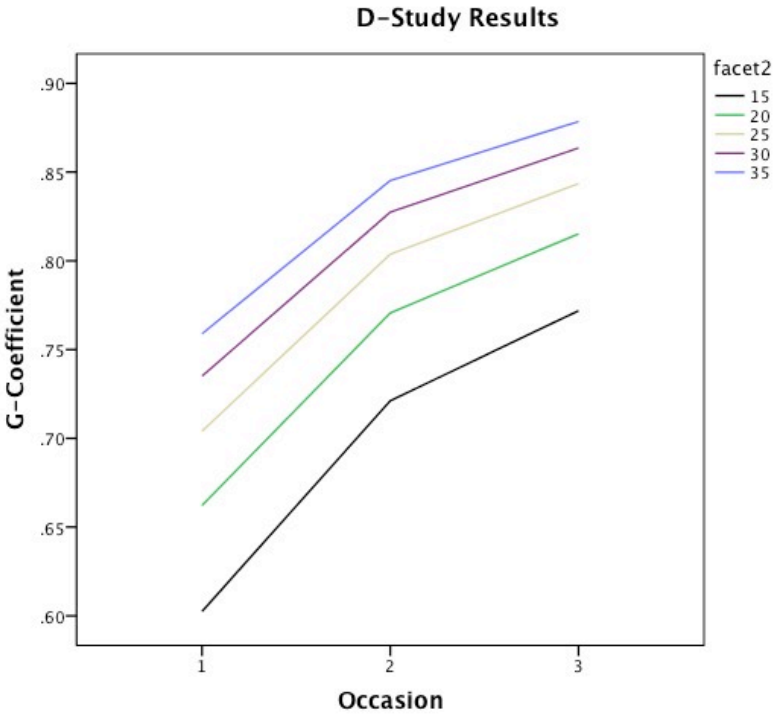
Forms	*n* Items				
	15	20	25	30	35
1	.016	.013	.011	.009	.008
2	.010	.008	.007	.006	.005
3	.008	.006	.005	.005	.004

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.641	.692	.726	.751	.770
2	.736	.780	.809	.830	.845
3	.774	.815	.841	.860	.874

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.624	.677	.713	.739	.759
2	.714	.761	.793	.816	.833
3	.750	.794	.824	.845	.860


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 6 Form 9

Variance Component	Proportion
Student	.082
Day	.002
Item	.069
Student * Item	.211
Student * Day	.005
Day * Item	.000
Student * Day * Item (residual)	.632

Error Variances:

Relative	*Absolute*
.007	.008

G-coefficients:

G	*Phi*
.741	.714

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.016	.013	.010	.009	.008
2	.011	.008	.007	.006	.005
3	.009	.007	.005	.005	.004

D-Study Item Relative Error Variances

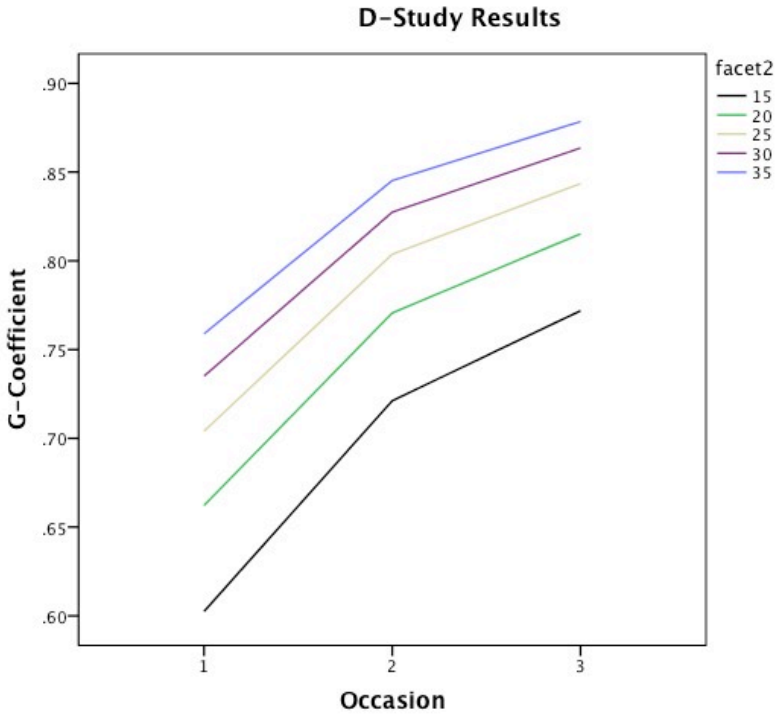
Forms	*n* Items				
	15	20	25	30	35
1	.015	.011	.009	.008	.007
2	.009	.007	.006	.005	.004
3	.007	.006	.005	.004	.003

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.574	.637	.681	.715	.741
2	.686	.741	.778	.805	.825
3	.734	.784	.817	.840	.858

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.550	.613	.658	.692	.718
2	.657	.714	.753	.781	.803
3	.702	.755	.791	.817	.836


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 6 Form 10

Variance Component	Proportion
Student	.082
Day	.000
Item	.060
Student * Item	.241
Student * Day	.002
Day * Item	.002
Student * Day * Item (residual)	.608

Error Variances:

Relative	*Absolute*
.008	.008

G-coefficients:

G	*Phi*
.729	.710

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.017	.013	.011	.009	.008
2	.011	.008	.007	.006	.005
3	.009	.007	.006	.005	.004

D-Study Item Relative Error Variances

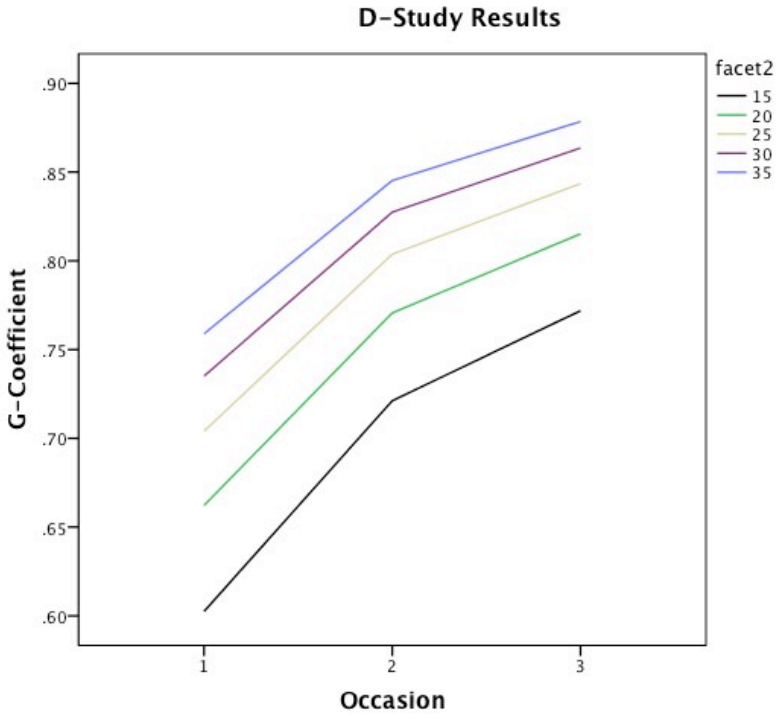
Forms	*n* Items				
	15	20	25	30	35
1	.016	.012	.010	.009	.008
2	.010	.008	.006	.005	.005
3	.008	.006	.005	.004	.004

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.565	.626	.669	.701	.726
2	.675	.729	.766	.792	.813
3	.721	.771	.805	.828	.846

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.549	.611	.656	.689	.715
2	.653	.710	.749	.777	.799
3	.696	.750	.786	.812	.831


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 6 Form n

Variance Component	Proportion
Student	.109
Day	.001
Item	.139
Student * Item	.174
Student * Day	.024
Day * Item	.002
Student * Day * Item (residual)	.551

Error Variances:

Relative	*Absolute*
.007	.009

G-coefficients:

G	*Phi*
.760	.723

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.018	.015	.013	.012	.011
2	.011	.009	.008	.007	.006
3	.009	.007	.006	.005	.005

D-Study Item Relative Error Variances

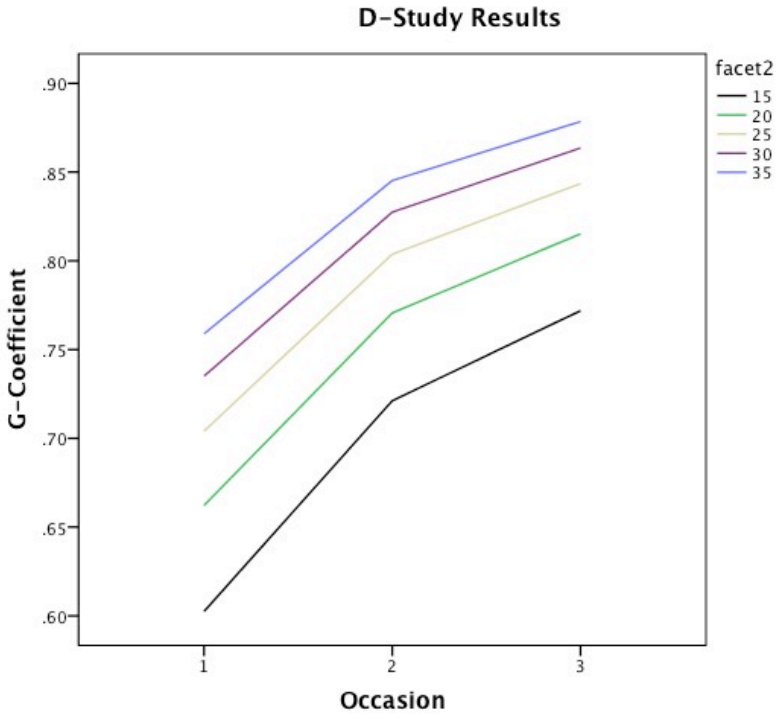
Forms	*n* Items				
	15	20	25	30	35
1	.016	.013	.011	.010	.010
2	.009	.007	.006	.006	.005
3	.007	.006	.005	.004	.004

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.602	.645	.673	.694	.710
2	.722	.760	.785	.802	.815
3	.774	.808	.830	.846	.857

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.569	.615	.647	.670	.687
2	.678	.723	.752	.773	.788
3	.725	.767	.795	.814	.829


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 7 Form 6

Variance Component	Proportion
Student	.074
Day	.000
Item	.095
Student * Item	.157
Student * Day	.010
Day * Item	.006
Student * Day * Item (residual)	.658

Error Variances:

Relative	*Absolute*
.007	.009

G-coefficients:

G	*Phi*
.716	.683

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.018	.014	.012	.010	.009
2	.011	.009	.007	.006	.005
3	.009	.007	.006	.005	.004

D-Study Item Relative Error Variances

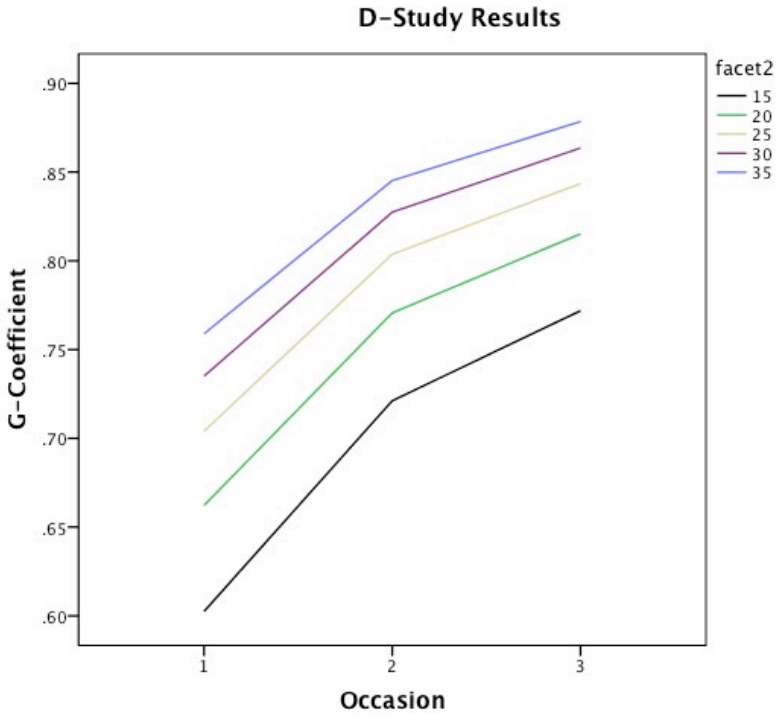
Forms	*n* Items				
	15	20	25	30	35
1	.016	.013	.011	.009	.008
2	.009	.007	.006	.005	.005
3	.007	.006	.005	.004	.004

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.534	.597	.634	.664	.688
2	.664	.716	.751	.777	.796
3	.722	.769	.800	.823	.839

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.509	.569	.612	.645	.670
2	.627	.683	.722	.751	.773
3	.679	.732	.768	.794	.814


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 7 Form 7

Variance Component	Proportion
Student	.071
Day	.001
Item	.053
Student * Item	.149
Student * Day	.007
Day * Item	.003
Student * Day * Item (residual)	.717

Error Variances:

Relative	*Absolute*
.007	.008

G-coefficients:

G	*Phi*
.711	.689

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.017	.013	.011	.010	.008
2	.010	.008	.007	.006	.005
3	.008	.006	.005	.004	.004

D-Study Item Relative Error Variances

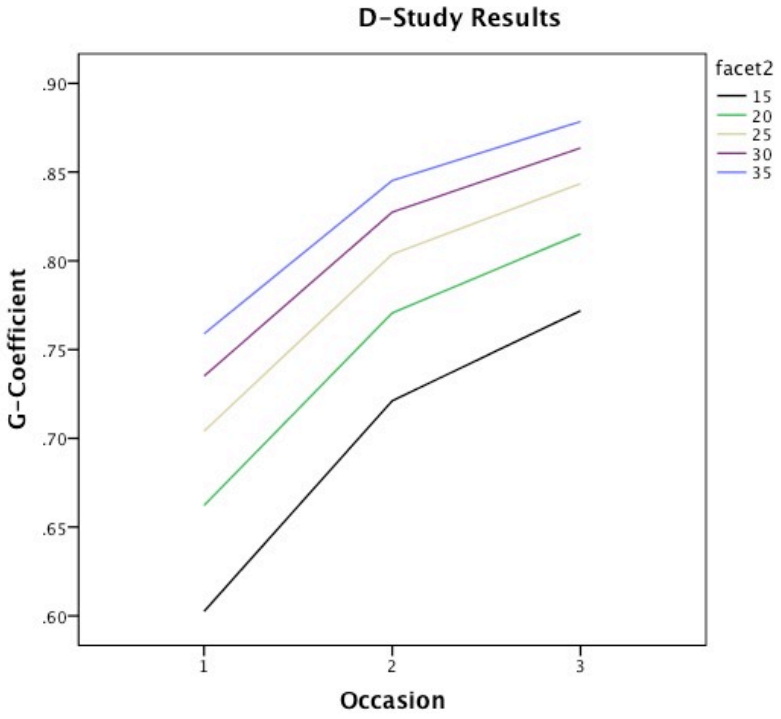
Forms	*n* Items				
	15	20	25	30	35
1	.016	.012	.010	.009	.008
2	.009	.007	.006	.005	.004
3	.007	.005	.004	.004	.003

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.523	.586	.631	.666	.693
2	.656	.711	.750	.778	.799
3	.716	.766	.800	.824	.842

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.506	.569	.614	.649	.676
2	.632	.689	.729	.759	.781
3	.689	.742	.778	.804	.824


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 7 Form 8

Variance Component	Proportion
Student	.075
Day	.000
Item	.068
Student * Item	.181
Student * Day	.000
Day * Item	.000
Student * Day * Item (residual)	.676

Error Variances:

Relative	*Absolute*
.006	.007

G-coefficients:

G	*Phi*
.742	.716

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.016	.012	.009	.008	.007
2	.010	.007	.006	.005	.004
3	.008	.006	.005	.004	.003

D-Study Item Relative Error Variances

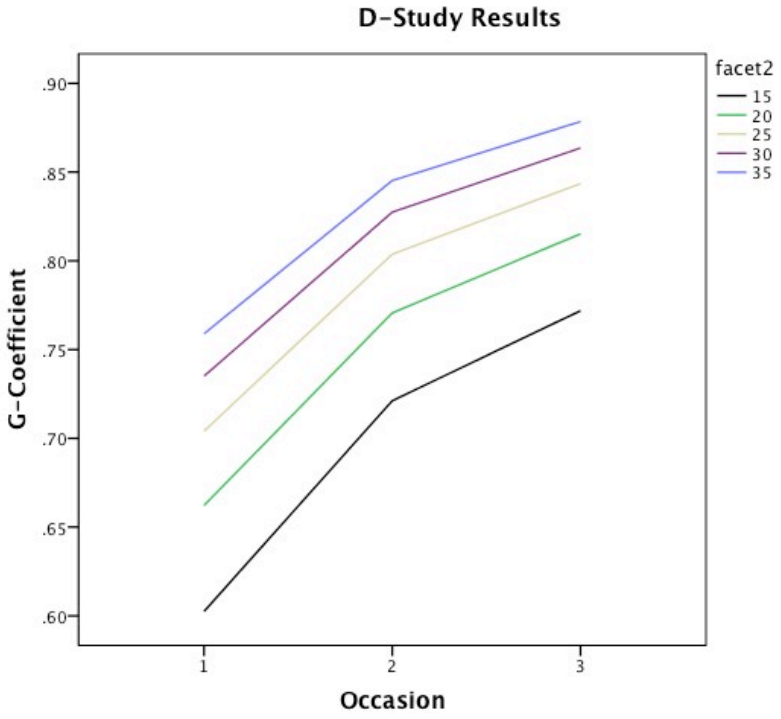
Forms	*n* Items				
	15	20	25	30	35
1	.014	.011	.009	.007	.006
2	.009	.006	.005	.004	.004
3	.007	.005	.004	.003	.003

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.566	.635	.685	.723	.753
2	.683	.742	.782	.812	.834
3	.734	.786	.821	.846	.865

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.546	.615	.666	.704	.735
2	.655	.716	.759	.790	.814
3	.701	.758	.796	.824	.845


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 7 Form 9

Variance Component	Proportion
Student	.050
Day	.001
Item	.087
Student * Item	.193
Student * Day	.010
Day * Item	.005
Student * Day * Item (residual)	.655

Error Variances:

Relative	*Absolute*
.008	.009

G-coefficients:

G	*Phi*
.616	.582

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.018	.014	.012	.011	.009
2	.012	.009	.007	.006	.006
3	.009	.007	.006	.005	.004

D-Study Item Relative Error Variances

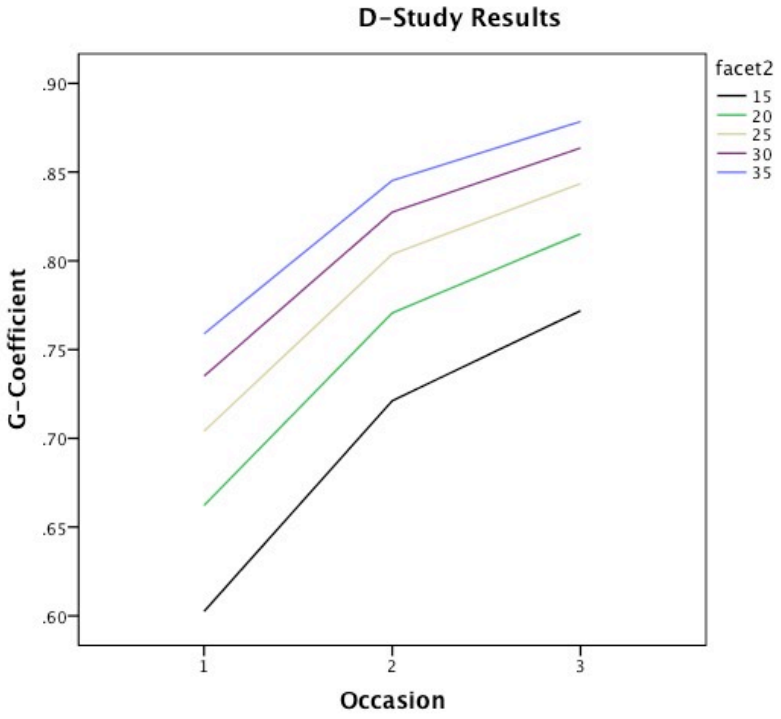
Forms	*n* Items				
	15	20	25	30	35
1	.017	.013	.011	.010	.009
2	.010	.008	.006	.006	.005
3	.008	.006	.005	.004	.004

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.428	.488	.532	.566	.593
2	.557	.616	.659	.690	.715
3	.619	.676	.716	.745	.768

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.405	.463	.508	.542	.570
2	.520	.582	.626	.660	.686
3	.575	.636	.679	.711	.736


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 7 Form 10

Variance Component	Proportion
Student	.091
Day	.000
Item	.071
Student * Item	.203
Student * Day	.000
Day * Item	.000
Student * Day * Item (residual)	.608

Error Variances:

Relative	*Absolute*
.007	.008

G-coefficients:

G	*Phi*
.773	.751

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.015	.011	.009	.008	.006
2	.010	.008	.006	.005	.004
3	.008	.006	.005	.004	.004

D-Study Item Relative Error Variances

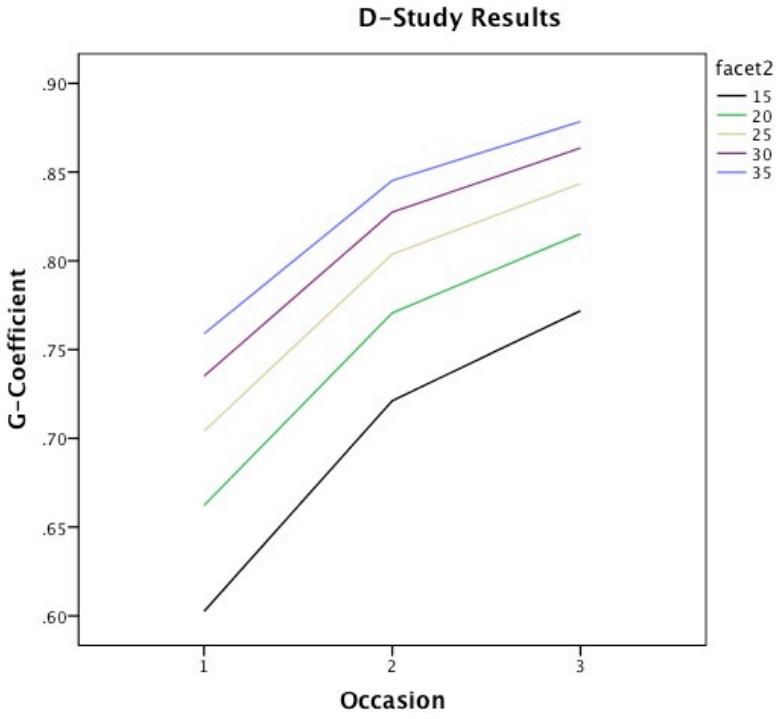
Forms	*n* Items				
	15	20	25	30	35
1	.014	.010	.008	.007	.006
2	.009	.007	.005	.004	.004
3	.007	.005	.004	.004	.003

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.620	.685	.731	.765	.792
2	.719	.773	.810	.837	.857
3	.760	.808	.840	.863	.881

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.601	.667	.715	.750	.778
2	.693	.751	.790	.819	.841
3	.731	.784	.819	.844	.864


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 7 Form n

Variance Component	Proportion
Student	.083
Day	.000
Item	.149
Student * Item	.165
Student * Day	.008
Day * Item	.002
Student * Day * Item (residual)	.593

Error Variances:

Relative	*Absolute*
.007	.008

G-coefficients:

G	*Phi*
.753	.705

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.016	.013	.011	.009	.008
2	.011	.008	.007	.006	.005
3	.009	.007	.006	.005	.004

D-Study Item Relative Error Variances

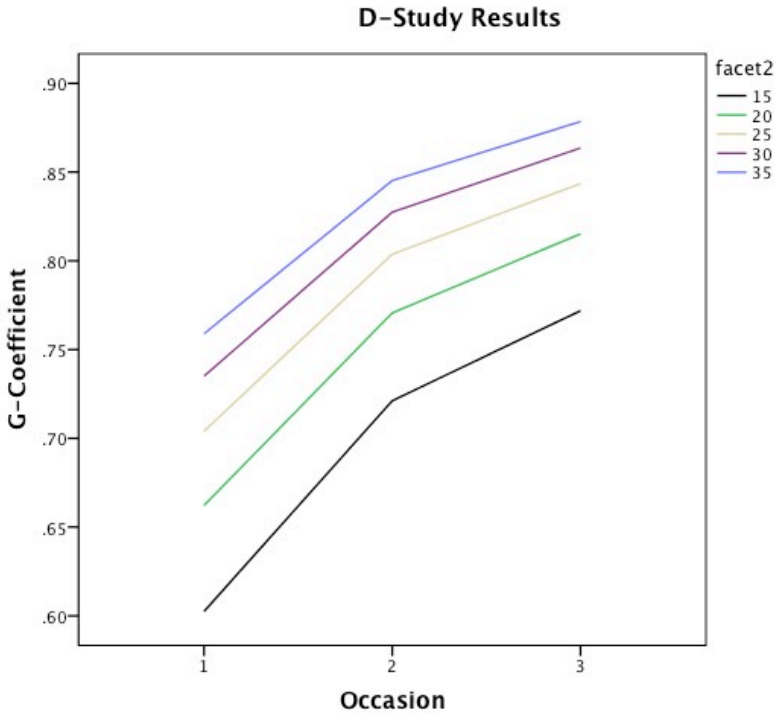
Forms	*n* Items				
	15	20	25	30	35
1	.014	.011	.009	.008	.007
2	.008	.007	.005	.005	.004
3	.006	.005	.004	.004	.003

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.586	.643	.684	.714	.737
2	.704	.753	.787	.810	.828
3	.755	.799	.828	.849	.864

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.547	.607	.651	.684	.709
2	.649	.705	.744	.772	.794
3	.692	.745	.781	.807	.827


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 8 Form 6

Variance Component	Proportion
Student	.055
Day	.000
Item	.047
Student * Item	.206
Student * Day	.000
Day * Item	.000
Student * Day * Item (residual)	.692

Error Variances:

Relative	*Absolute*
.007	.007

G-coefficients:

G	*Phi*
.666	.648

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.016	.012	.009	.008	.007
2	.010	.007	.006	.005	.004
3	.008	.006	.005	.004	.003

D-Study Item Relative Error Variances

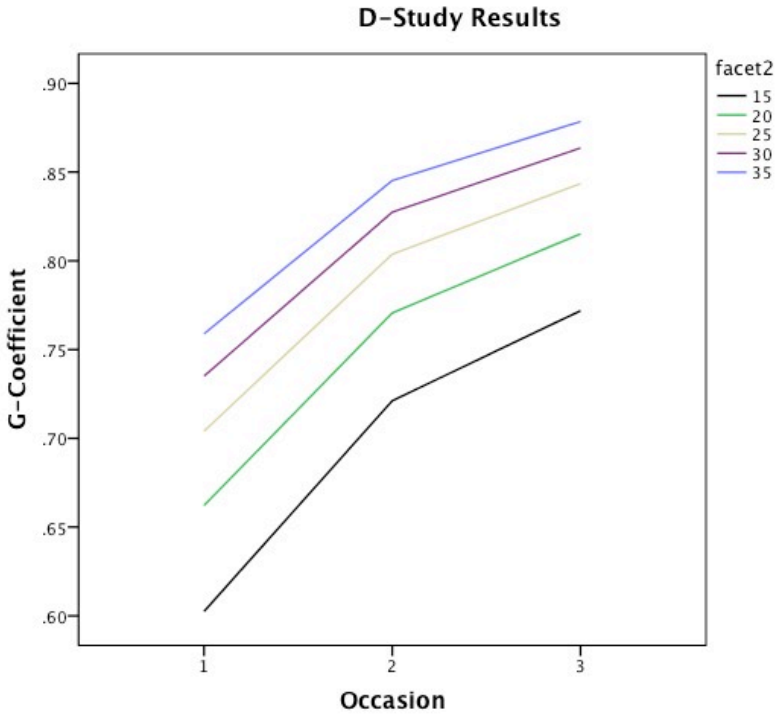
Forms	*n* Items				
	15	20	25	30	35
1	.015	.011	.009	.007	.006
2	.009	.007	.006	.005	.004
3	.007	.005	.004	.004	.003

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.479	.551	.605	.648	.682
2	.600	.666	.714	.750	.778
3	.654	.716	.759	.791	.815

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.467	.539	.593	.636	.671
2	.580	.648	.697	.734	.763
3	.631	.695	.740	.774	.800


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 8 Form 7

Variance Component	Proportion
Student	.069
Day	.000
Item	.036
Student * Item	.209
Student * Day	.000
Day * Item	.000
Student * Day * Item (residual)	.686

Error Variances:

Relative	*Absolute*
.007	.007

G-coefficients:

G	*Phi*
.715	.702

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.015	.012	.009	.008	.007
2	.010	.007	.006	.005	.004
3	.008	.006	.005	.004	.003

D-Study Item Relative Error Variances

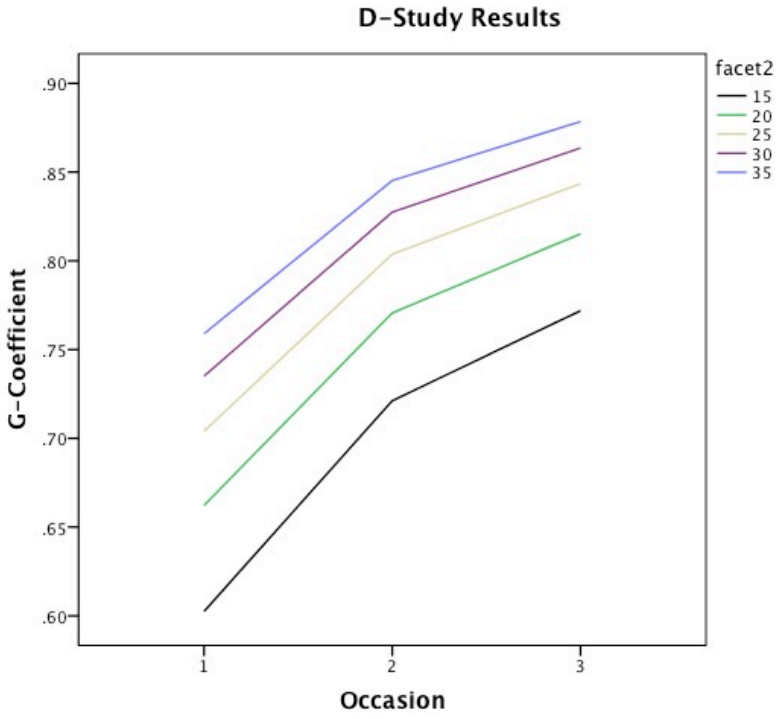
Forms	*n* Items				
	15	20	25	30	35
1	.015	.011	.009	.007	.006
2	.009	.007	.005	.005	.004
3	.007	.005	.004	.004	.003

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.538	.608	.660	.699	.731
2	.653	.715	.759	.790	.815
3	.704	.760	.799	.826	.847

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.528	.598	.651	.691	.723
2	.639	.702	.747	.780	.805
3	.687	.746	.786	.815	.837


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 8 Form 8

Variance Component	Proportion
Student	.081
Day	.000
Item	.045
Student * Item	.167
Student * Day	.007
Day * Item	.005
Student * Day * Item (residual)	.695

Error Variances:

Relative	*Absolute*
.007	.008

G-coefficients:

G	*Phi*
.736	.720

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.017	.013	.011	.009	.008
2	.010	.008	.006	.006	.005
3	.008	.006	.005	.004	.004

D-Study Item Relative Error Variances

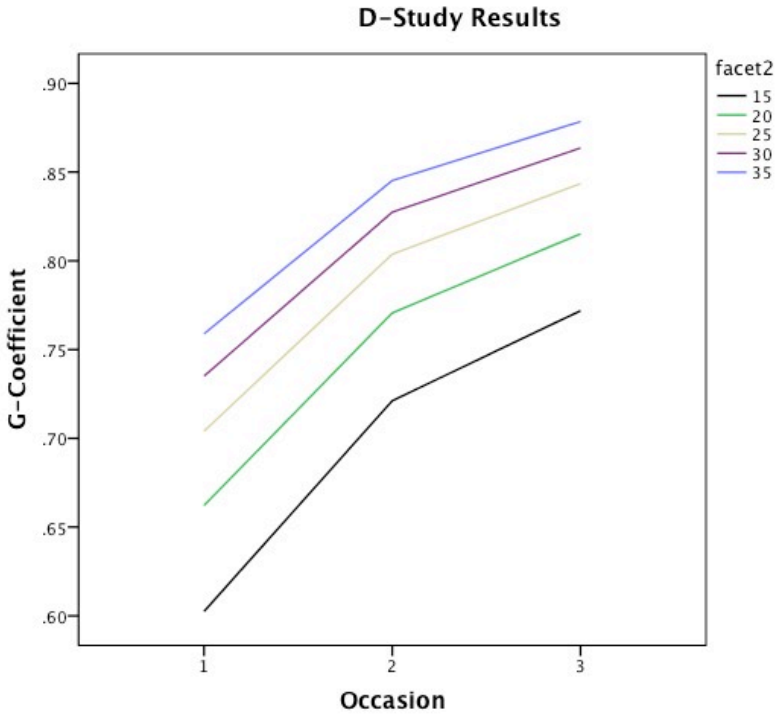
Forms	*n* Items				
	15	20	25	30	35
1	.016	.012	.010	.009	.008
2	.009	.007	.006	.005	.005
3	.007	.006	.005	.004	.003

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.558	.620	.663	.696	.722
2	.683	.736	.772	.798	.818
3	.738	.785	.817	.839	.856

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.545	.607	.651	.685	.711
2	.665	.720	.758	.785	.806
3	.717	.767	.801	.825	.844


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 8 Form 9

Variance Component	Proportion
Student	.069
Day	.000
Item	.030
Student * Item	.151
Student * Day	.011
Day * Item	.002
Student * Day * Item (residual)	.737

Error Variances:

Relative	*Absolute*
.008	.008

G-coefficients:

G	*Phi*
.687	.676

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.018	.014	.012	.010	.009
2	.011	.008	.007	.006	.005
3	.008	.006	.005	.004	.004

D-Study Item Relative Error Variances

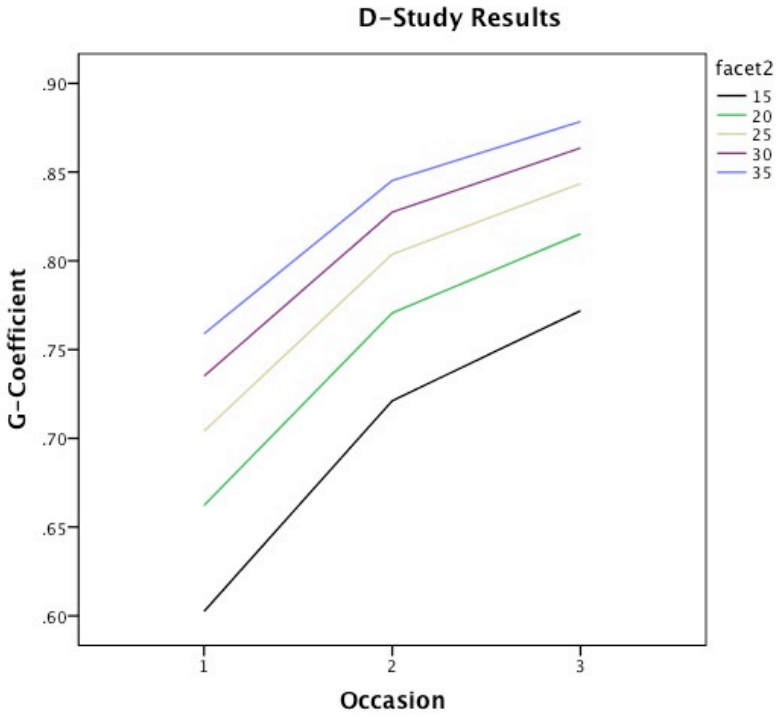
Forms	*n* Items				
	15	20	25	30	35
1	.018	.014	.012	.010	.009
2	.011	.008	.007	.006	.005
3	.008	.006	.005	.004	.004

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.496	.555	.598	.630	.655
2	.633	.687	.724	.752	.773
3	.696	.746	.780	.804	.822

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.488	.548	.591	.624	.650
2	.621	.676	.715	.743	.765
3	.682	.734	.769	.794	.813


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 8 Form 10

Variance Component	Proportion
Student	.085
Day	.001
Item	.028
Student * Item	.181
Student * Day	.011
Day * Item	.006
Student * Day * Item (residual)	.690

Error Variances:

Relative	*Absolute*
.008	.008

G-coefficients:

G	*Phi*
.729	.717

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.018	.014	.012	.010	.009
2	.011	.008	.007	.006	.005
3	.008	.006	.005	.005	.004

D-Study Item Relative Error Variances

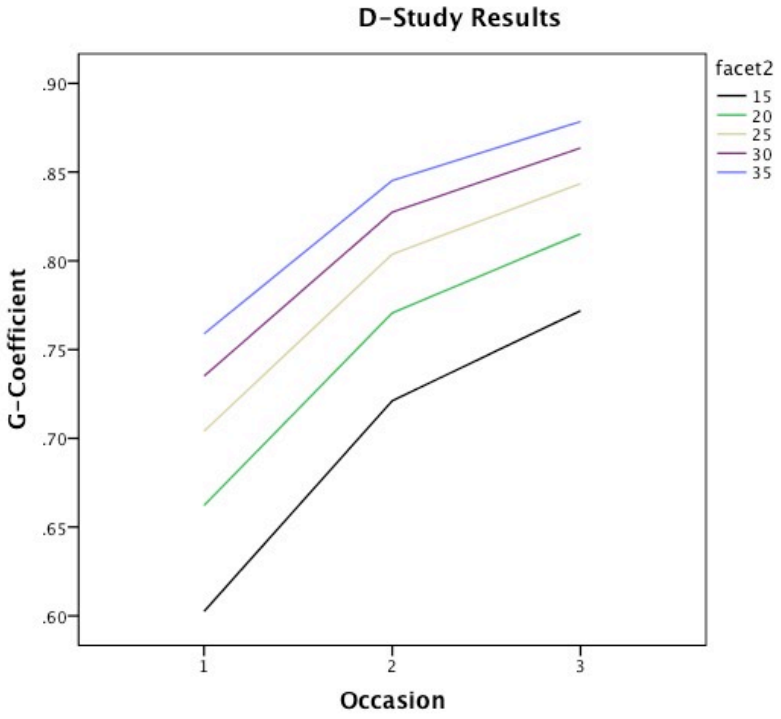
Forms	*n* Items				
	15	20	25	30	35
1	.017	.014	.011	.010	.009
2	.010	.008	.007	.006	.005
3	.008	.006	.005	.004	.004

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.553	.611	.651	.682	.705
2	.678	.729	.763	.788	.807
3	.733	.779	.809	.831	.847

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.543	.601	.642	.673	.696
2	.665	.717	.753	.779	.798
3	.719	.767	.799	.822	.839


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Grade 8 Form n

Variance Component	Proportion
Student	.096
Day	.000
Item	.039
Student * Item	.163
Student * Day	.006
Day * Item	.001
Student * Day * Item (residual)	.696

Error Variances:

Relative	*Absolute*
.007	.008

G-coefficients:

G	*Phi*
.771	.759

D-Study Item Absolute Error Variances

Forms	*n* Items				
	15	20	25	30	35
1	.016	.013	.010	.009	.008
2	.010	.008	.006	.005	.005
3	.007	.006	.005	.004	.004

D-Study Item Relative Error Variances

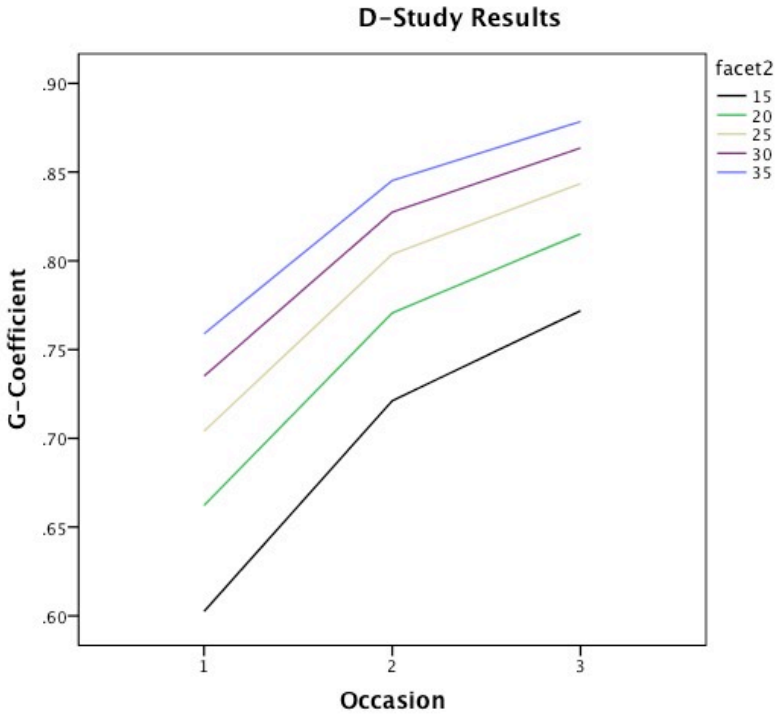
Forms	*n* Items				
	15	20	25	30	35
1	.016	.012	.010	.009	.008
2	.009	.007	.006	.005	.004
3	.007	.005	.004	.004	.003

D-Study G-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.602	.662	.704	.735	.759
2	.721	.771	.804	.827	.845
3	.772	.815	.844	.864	.879

D-Study Phi-Coefficients

Forms	*n* Items				
	15	20	25	30	35
1	.592	.653	.696	.728	.752
2	.707	.759	.793	.818	.837
3	.756	.802	.832	.853	.869


D-Study Results
G-Coefficient
.90
.85
.80
.75
.70
.65
.60
1
2
3
Occasion
facet2
15
20
25
30
35

Form G-Theory Analysis: Grade 6

Variance Component	Proportion
Student	.382
Day	.000
Valid	.337
Forms:Persons	.100
Student*Day	.034
Student*Valid	.000
Day*Valid	.007
Residual	.139

Error Variances:

Relative	*Absolute*
1.425	1.776

G-coefficients:

G	*Phi*
.847	.816

D-Study Item Absolute Error Variances

Forms	*n* Occassions		
	1	2	3
1	8.051	4.432	3.226
2	7.549	4.004	2.822
3	7.382	3.861	2.687

Note. Analysis assumes all students responded to all items on a 20 item test

D-Study Item Relative Error Variances

Forms	*n* Occassions		
	1	2	3
1	.950	.886	.862
2	.532	.495	.483
3	.389	.365	.357

Note. Analysis assumes all students responded to all items on a 20 item test

D-Study G-Coefficients

Forms	*n* Occassions		
	1	2	3
1	.892	.899	.901
2	.937	.941	.942
3	.953	.956	.957

Note. Analysis assumes all students responded to all items on a 20 item test

D-Study Phi-Coefficients

Forms	*n* Occassions		
	1	2	3
1	.495	.640	.709
2	.511	.663	.736
3	.516	.671	.746

Note. Analysis assumes all students responded to all items on a 20 item test

Form G-Theory Analysis: Grade 7

Variance Component	Proportion
Student	.374
Day	.004
Valid	.379
Forms:Persons	.037
Student*Day	.000
Student*Valid	.000
Day*Valid	.000
Residual	.205

Error Variances:

Relative	*Absolute*
0.363	0.724

G-coefficients:

G	*Phi*
.946	.898

D-Study Item Absolute Error Variances

Forms	*n* Occassions		
	1	2	3
1	6.760	3.435	2.327
2	6.632	3.352	2.258
3	6.590	3.324	2.235

Note. Analysis assumes all students responded to all items on a 20 item test

D-Study Item Relative Error Variances

Forms	*n* Occassions		
	1	2	3
1	.212	.124	.095
2	.122	.078	.063
3	.092	.062	.053

Note. Analysis assumes all students responded to all items on a 20 item test

D-Study G-Coefficients

Forms	*n* Occassions		
	1	2	3
1	.968	.981	.985
2	.981	.988	.990
3	.986	.990	.992

Note. Analysis assumes all students responded to all items on a 20 item test

D-Study Phi-Coefficients

Forms	*n* Occassions		
	1	2	3
1	.486	.650	.733
2	.490	.656	.739
3	.492	.658	.741

Note. Analysis assumes all students responded to all items on a 20 item test

Form G-Theory Analysis: Grade 8

Variance Component	Proportion
Student	.378
Day	.000
Valid	.282
Forms:Persons	.078
Student*Day	.032
Student*Valid	.028
Day*Valid	.000
Residual	.203

Error Variances:

Relative	*Absolute*
0.980	1.215

G-coefficients:

G	*Phi*
.865	.838

D-Study Item Absolute Error Variances

Forms	*n* Occassions		
	1	2	3
1	5.926	3.259	2.370
2	5.578	2.953	2.078
3	5.461	2.851	1.981

Note. Analysis assumes all students responded to all items on a 20 item test

D-Study Item Relative Error Variances

Forms	*n* Occassions		
	1	2	3
1	1.227	.910	.804
2	.879	.604	.512
3	.763	.502	.415

Note. Analysis assumes all students responded to all items on a 20 item test

D-Study G-Coefficients

Forms	*n* Occassions		
	1	2	3
1	.837	.874	.887
2	.877	.912	.925
3	.892	.926	.938

Note. Analysis assumes all students responded to all items on a 20 item test

D-Study Phi-Coefficients

Forms	*n* Occassions		
	1	2	3
1	.515	.659	.726
2	.530	.680	.752
3	.535	.688	.760

Note. Analysis assumes all students responded to all items on a 20 item test