Developing a Large-scale Assessment Using Components of Evidence-centered Design: Did it Work?[1]

Claudia Flowers, UNC Charlotte

Charlene Turner, Bill Herrera, Liz Towles-Reeves, edCount, LLC

Martha Thurlow, University of Minnesota

Anne Davidson, Smarter Balanced

and

Sarah Hagge, Minnesota Department of Health

A paper presented at the annual National Council on Measurement in Education, Chicago, IL, 2015.

[1] Disclaimer: The contents of this paper were developed as part of the National Center and State Collaborative under a grant from the U.S. Department of Education (PR/Award # H373X100002), Project Officer, Susan.Weigert@ed.gov. However, the content do not necessarily represent the policy of the U.S. Department of Education and no assumption of endorsement by the Federal government should be made.

Abstract

Evidence-centered design (ECD) provides a systematic framework for designing assessments in terms of evidentiary arguments. The purpose of this paper is to report the results of implementing components of the ECD approach in the development of an alternate assessment for students with significant cognitive disabilities. Examples from the 3rd grade mathematics assessment are presented to portray the components of the ECD approach used by NCSC. Results from a field test of over 6000 students provided empirical evidence that supported the test's conceptual assessment framework. Implications for using ECD are presented.

Keywords

Evidence-centered Design, Large-scale Assessment, Students with Significant Cognitive Disabilities

Developing a Large-scale Assessment Using Evidence-centered Design: Did it Work?

Evidence-centered design (ECD), first proposed by Mislevy, Steinberg, and Almond (2003), conceptualized validity as an argument and chain of reasoning specified to students, evidence, and task models described by Messick (1994). The ECD approach is structured as a sequence of test development layers that include (a) domain analysis, (b) domain modeling, (c) conceptual assessment framework development, (d) assessment implementation, and (e) assessment delivery (Mislevy & Riconscente, 2005; Mislevy, Steinberg, & Almond, 2003). Since the original introduction of ECD, the principles, patterns, examples, common language and knowledge representations for designing, implementing and delivering educational assessment using the processes of ECD have been further elaborated (Mislevy & Haertel, 2006).

The National Center and State Collaborative (NCSC) used components of ECD for developing a multi-state alternate assessment based on alternate achievement standards (AA-AAS) for students with the most significant cognitive disabilities (SWSCD). Over the course of AA-AAS development, NCSC developed a conceptual model for systematically varying item complexity across and within content standards and domains that incorporated the interaction between content aligned to the Common Core State Standards (CCSS), tasks, characteristics of SWSCD, and how SWSCD demonstrate what they know and can do. The guiding principle for the AA-AAS development was to create an AA-AAS, for 3rd to 8th and 11th grade in mathematics and English language arts, that: (a) was accessible to all students, (b) supported the score inferences, and (c) collected evidence to examine the interpretive argument.

The purpose of this study was to examine the ECD approach and empirically evaluate the theory underlying the constructs and the psychometric properties of the assessment items. The following questions were examined: (a) How were components of the ECD layers used to design AA-AAS? and (b) To what extent does the pattern of the item difficulty levels match the item complexity levels reflected in the item design?

Description of the Implementation of the Layers of ECD

Using the layers of ECD, a description of the implementation of ECD is presented in the development of mathematics AA-AAS. An example from 3rd grade mathematics AA-AAS development is included for illustration purposes. Before presenting the procedures used to implement the components of the ECD framework, a description of the student population and the need for developing a wide range of complexity into the items are provided.

Student Population

Students who participate in AA-AAS have been determined by an IEP team to meet criteria that indicate they should participate in an AA-AAS because they have significant cognitive disabilities. To establish an understanding of the target student population for the AA-AAS, researchers partnered with 18 states in 2012 to gather and analyze Learner Characteristics Inventory data (LCI; Kearns, Kleinert, Kleinert, & Towles-Reeves, 2006). Results for 48,669 students indicated that the majority of the students participating in AA-AAS were between the ages of 6 and 18 years old (Towles-Reeves, Kearns, Flowers, Hart, Kerbel, Kleinert, Quenemoen, & Thurlow, 2012). On average, intellectual disability was the most frequently reported disability category (56%), followed by autism spectrum disorder (22%) and multiple disabilities (9%). The majority of the students (69%) used symbolic language to communicate, with 18% of the students used intentional communication but not at a symbolic level (i.e., emerging symbolic), and 10% of the students communicated through cries, facial expressions, or change in muscle tone, but showed no clear use of objects/textures, regularized gestures, pictures, or signs in communication (pre-symbolic). Regarding receptive communication, about half of the students (49%) could independently follow 1 or 2 step directions, and 37% of the students required additional cues to follow 1 or 2 step directions. Most of the students (65%) could read written text or braille; 22% of the students could read fluently with basic, literal understanding.

Domain Analysis

Domain analysis involves determining the specific content to be included in the assessment. All AA-AAS must be aligned to grade-level content standards (U.S. Department of Education, 2004), a requirement that challenges all AA-AAS developers. Consistent with the National Research Council's assessment triangle (Pellegrino, Chudowsky, & Glaser, 2001), NCSC carefully considered a "model of how students represent knowledge and develop competence in the subject domain" (p. 2) in developing tasks and ways to interpret performance. NCSC identified meaningful academic content and ways to assess the content that addressed the range of characteristics and needs evidenced by the heterogeneous group of students who participate in an AA-AAS. Assessment tasks/items were created to capture student performance at different levels of proficiency against grade-level standards. At the same time, NCSC incorporated other important aspects of item design such as depth of knowledge, Lexile levels (reading level), and degree of scaffolding and support.

The ECD approach to assessment item development addressed these complexities by clarifying relationships and identifying key decision points. Explicit principles of Universal Design for Learning (UDL; Dolan, Rose, Burling, Harms, & Way, 2007) were included in the design of assessment tasks and features based on these principles. The resulting Task Templates and Design Patterns (tools built into the ECD process that were precursors to item development) served as the mechanism by which varying levels of content difficulty were implemented in the family of assessment items measuring a particular aspect of the core content identified for the NCSC AA-AAS.

Mathematics content domain. To reduce the breadth of CCSS, NCSC was guided by the five goals identified by Kearns et al. (2010) as essential to ensuring that SWSCD become college and career ready: (1) communication, (2) fluency of the use of content, (3) age appropriateness, (4) independence, and (5) support systems. Next, NCSC used a Learning Progressions Framework (LPF) to present a broad description of the essential content and general sequencing for student learning and skill development – the pathway that typical peers may take grade by grade (Hess, 2010; National Resource Council, 2010). The LPF was developed using existing research on the sequence of learning that occurs in typically developing students. The LPF provided a mechanism for understanding the educational logic and pathway to help move SWSCD toward the CCSS.

To maintain alignment to the CCSS and select targeted academic content as described in the LPF, Core Content Connectors (CCCs) were developed that bridged both CCSS and LPF. NCSC developed the CCCs to identify the most salient core academic content in mathematics found in both the CCSS and the LPF. The CCCs illustrate the necessary knowledge and skills SWSCD need to reach the learning targets or critical big ideas within the LPF and the CCSS. In total, 482 mathematics CCCs were developed, with 38 to 58 CCCs per grade. The CCCs retained the grade-level content focus of these two resources; they were not extended standards (i.e., reduction in complexity). This approach required item developers to focus on how to build a range of complexity into the assessment while maintaining the grade-level content standards.

Central to all decisions about the development of the NCSC AA-AAS was the involvement of special educators. They provided input on the learning processes of SWSCD, and the importance of creating AA-AAS that would allow all students to demonstrate what they know and can do in the grade-level content. In addition to creating the CCCs, *essential understandings* were developed that described both the concrete and symbolic representational understandings necessary to engage with the academic content for those students who had limited use of symbols.

After the CCCs were developed across all grades (i.e., grades K-12) and all content areas, NCSC selected 10 CCCs as prioritized content for inclusion in the mathematics AA-AAS at a particular grade-level. The decision to target 10 CCCs for prioritization was based on research that examined SWSCD mastery of content that was linked to the CCSS (Browder, Flowers, Saunders, Bethune, & Spooner, 2013). Browder et al. found that given two weeks of evidence-centered instruction, 74% of students, across all lessons, made some progress. It was estimated that if teachers were to focus instruction on 10 standards, this would require more than 2 weeks of instruction for each standard to achieve mastery and generalization. Considering the curriculum demands on teachers of SWSCD, the need to reteach many of the knowledge, skills, and abilities, and the need to test at least two academic content areas, 10 CCCs appeared to be a reasonable starting point for focusing the NCSC AA-AAS.

The mathematics prioritization for selecting the 10 CCCs was informed by how mathematics develops in the CCSS. The goal of the prioritization was to ensure the mathematics assessment supported instruction of grade-specific skills and concepts, as well as higher expectations for SWSCD.

NCSC began the prioritization process by developing four mathematics claims. The claims identified grade-level proficiency, showed how knowledge and skills are built over time, identified evidence, and indicated the kinds of situations – the tasks or items – that would give students the optimal opportunity to produce the desired evidence. Math claims were developed based, in part, on those that had been developed by the Race-to-the-Top general assessment consortia (Smarter Balanced and PARCC) with NCSC mapping the prioritized math CCCs to the claims to be sure there would be sufficient evidence to support the claims.

The mathematics claims were that SWSCD can: (a) perform mathematical procedures with accuracy; (b) demonstrate or explain the application of mathematical concepts; (c) interpret or represent quantitative relationships using mathematical tools, such as manipulatives, models, rules, or symbols; and (d) make sense of problems based on real-world scenarios, choose an appropriate strategy, and apply mathematics to find a solution. An example of a CCC with its associated CCSS, *essential understanding*, rationale for inclusion in AA-AAS, and claim is provided in Table 1.

Conceptual mathematical model for SWSCD. Browder and Lee (2014) developed a conceptual model for mathematics instruction for SWSCD that guided the design choices. The model focuses on teaching students the mathematical content of their assigned grade and chronological age, with prioritization of the content and supports to compensate for not yet mastered prerequisites. This grade-aligned mathematical model, which includes ongoing promotion of numeracy skills, provides SWSCD the greatest access to general curriculum content. For this model to work, students should receive intensive early instruction to gain as many early mathematics skills as possible. After the early grades, students should focus on the standards of their grade level, building these early skills through new applications that are grade and age appropriate. To make this content accessible, Browder and Lee recommended using problems that provide a familiar context for students through real life applications and other high interest themes. The math problems themselves may need to have simplified language and be presented with read-alouds. Students may need to learn and demonstrate the mathematical skill through a step-by-step process, and students also may need supports such as calculators and manipulatives to augment underdeveloped skills. Through this approach, students can not only participate in general education mathematics, but also build a broader range of skills needed to function in the 21st century.

[Insert Table 1 about here]

Domain Modeling

Domain modeling entails creation and documentation of a high-level description of the components of the assessment argument. Using Principled Assessment Designs for Inquiry (PADI: see Baxter & Mislevy, 2005 for details), Design Patterns were created that identified attributes that address the necessary elements of an assessment argument. Each Design Pattern detailed three essential elements around which all assessments revolve: (a) the student’s knowledge, skills, and abilities (KSAs) for making inferences, (b) the salient characteristics of what students say, do, or make that would provide evidence about acquisition of the Focal KSAs, and (c) features of the task environment that are needed to evoke the desired evidence (Haertel & Cameto, 2013). These three elements served as the building blocks that the assessment designers used for an intentional process of task design, with a goal being to support a coherent assessment argument.

Item complexity was designed across two dimensions: (a) cognitive complexity of the academic content, and (b) degree of scaffolding applied to support student performance. Variable features (i.e., item features that can be adjusted in order to increase or decrease the difficulty of items) allowed item developers to vary the difficulty and mitigate construct-irrelevant variance (Haertel & Cameto, 2013). For a full description of the 19 variable features, see Mislevy et al. (2003). An example of some of the attributes addressed in a NCSC Design Pattern is displayed in Table 2.

[Insert Table 2 about here]

Conceptual Assessment Framework

The NCSC conceptual assessment framework specified in detail the KSAs to be assessed, the evidence that needed to be collected, and the features of the tasks that would elicit the evidence. Also identified were non-targeted KSAs, which although required for successful performance on an item, were not the intended target of the assessment. By identifying non-targeted KSAs, construct-irrelevant variance was minimized and accessibility maximized. Finally, the psychometric model and evaluative decision rules for task scoring were considered and assessment task features were defined and carefully aligned with the targeted and non-targeted KSAs.

Using information from the Design Patterns, Task Templates were created to demonstrate how key attributes could be incorporated into item development. NCSC wanted to develop items that would reflect a range of performance while ensuring accessibility for all students. Toward this end, items were designed across four complexity levels for each prioritized KSA. The four items of graduated complexity within a specific area of the content formed an *item family,* and the levels were referred to as tiers. Tier 1 items focused on *essential understanding* of the CCC, had the least complex content, and incorporated greater use of non-construct relevant scaffolds. Tier 2 to 4 items assessed the focal KSA, with varying complexity of content and use of scaffolds (i.e., complexity increases, with the tier 2 item being less complex, to the tier 4 item being most complex and closest to grade-level expectation with few, if any scaffolds). The Task Templates also served to illustrate to item writers the boundaries for developing an item and included the attributes that were developed in the Design Pattern. An example of how the Task Template uses attributes is shown in Table 3. Although potential observations at each of the tier levels are included in the NCSC Task Templates, they are not included in this paper for test security reasons. A mathematics practice item at each tier is shown in Figure 1.

[Insert Table 3 and Figure 1 about here]

Assessment Implementation

Pilot test. In fall 2014, NCSC conducted a pilot test of its mathematics AA-AAS for grades 3-8 and 11. Because the assessments were administered at the beginning of the school year, the student sample was tested on the previous grade's test; that is, for example, the grade 3 test was administered to students in grade 4. Nineteen states participated and more than 6000 students were administered the NCSC AA-AAS. The student sampling plan was designed to be representative of the 19 states that participated. Approximately 64% of the students were male and 60% white. The learner characteristics were similar to those reported earlier.

Four mathematics forms were available for each grade, with a total of 40 items per form. All forms within a grade covered the same proportion of content, with a tier distribution of: (a) 20% tier 1, (b) 35% tier 2, (c) 35% tier 3, and (d) 10% tier 4. The intent was to have all forms covering the same content at the same difficulty level. The assessment was divided into two sessions, with the first session containing 20 common (anchor) items for the four forms. Within a grade, test forms were randomly spiraled by school or classroom.

Rasch calibrations. Rasch IRT concurrent calibration, implemented using WINSTEPS, was used to place items on the same scale. Session 1 items were used to anchor the Session 2 items to a common scale. Model-data fit was examined using Mean Square (MS) infit and MS outfit statistics. To assess dimensionality, a principal component analysis of the residual variation evaluated the unidimensionality assumption.

Results. The number of students, mean raw scores, mean difficulty value (*p*-value), and Cronbach's alpha by grade and form are reported in Table 4. The number of students per form ranged from 77 to 210. The mean raw score ranged from 43% to 53% of total points per form. Cronbach alpha coefficients ranged from .66 to .88.

[Insert Table 4 about here]

The average Rasch item parameter (IP) measures (i.e., logit) by grade and tier are reported in Table 5. Recall that the concurrent calibration placed all items on the same scale. In all cases the average IP followed the pattern expected as designed; that is, on average the items were more difficult as the tier level increased. There was one exception in the 5th grade at tiers 3 and 4 (in bold), but this anomaly could be attributed to error. A graphic display of the data is provided in Figure 2.

[Insert Table 5 and Figure 2 about here]

Results of the two-way ANOVA with the Rasch IP as the dependent variable and tier and grade as the independent variables indicated there was a statistically significant tier effect ($F_{3,672}$=245.04, p<.001, partial eta squared=.52), but not a statistically significant grade ($F_{6,672}$=.20, p=.94, partial eta squared=.002), or interaction effect ($F_{18,672}$=1.50, p=.08, partial eta squared=.039). Post hoc analyses that examined differences between the tiers are reported in Table 6. There were statistically significant differences among tier 1 and tiers 2 to 4, with very large effect sizes (Cohen's *d*) ranging from 1.67 to 3.66. For tier 2 compared to tiers 3 and 4, there were statistically significant differences between the tier 3 items for four of the seven grades and for tier 4 there were differences in six of the seven grades. When comparing tier 3 items to tier 4, there were no statistically significant differences.

[Insert Table 6 about here]

To control for possible content effects, analyses were conducted at the family level (the four tier items that target the same KSA). The percentage of items that performed as expected (e.g., tier 1 IP is less than tier 2 IP), difference in the Rasch IP (e.g., IP difference = T1-T2), and the number of paired items within families are shown in Table 7. Similar to the previous analyses, tier 1 items almost always were the easier items when compared to the other tiers, with 88% to 100% of the items being the easiest within the item family. For tiers 2 to 4, the items generally performed as expected with a few exceptions. In grades 3 and 5, when examining tier 3 verses tier 4, the tier 4 items within the family tended to be easier (bolded in Table).

[Insert Table 7 about here]

A possible explanation for the higher tier items (i.e., tiers 3 and 4) not creating more separation in the difficulty level may be the students' lack of opportunity to learn and the lack of student and teacher motivation. Although NCSC has developed curriculum materials that are aligned to the AA-AAS, most teachers reported in the end-of-test survey that their students had not received instruction in the more academically challenging math content. To examine student guessing in the higher tiers, an index was created using the *p*-value and subtracting the chance level. For example, for a tier 2 item that had a *p*-value of .5, given that the items have three-response option resulting in a random selection of the correct answer .33 proportion of the time, the guessing index would be .17 (index = .50 - .33). The indices for each grade at each tier are reported in Table 8. On average, most of the tier 4 items were close to chance (i.e., close to 0.0). Tiers 1 and 2 items had somewhat larger indices, suggesting that students are not guess as much on these items.

[Insert Table 8 about here]

NCSC examined dimensionality of the data using principal component analysis of residual variation, as produced by WINSTEPS. Results of the analyses are reported in Table 9 with the first column of the table indicating the grade. The second column of the table indicates the number of items, and the remaining three columns indicate the percent of variation associated with each dimension. As expected, the first dimension associated with the Rasch model accounts for the majority of score variation.

[Insert Table 9 about here]

Assessment Delivery

The final ECD layer is assessment delivery, which coordinates the interactions of students to tasks and scoring. Four principal processes were used: (a) item selection process, (b) presentation process, (c) evidence identification process, and (d) evidence accumulation.

Test blueprints were designed to determine content coverage and complexity of content. Item field test data and small scale item tryouts provided NCSC insight into how students interact with the items and which items could be synthesized into the evidence about the claims.

Over the course of item and test development, NCSC has progressed toward the goal of developing a defensible and innovative online summative assessment program for SWSCD. The online platform provided various assessment features to support student access to the test and incorporated the elements of UDL (Dolan et al., 2007). Assessment features were either built into the NCSC Assessment System or were typically available on a computer. Assessment features could be enabled by the student or test administrator (TA) at the time of the testing. The test was designed to have all items and response options read to the student, by either the Audio Player or the TA.

The NCSC *Test Administration Manual* (TAM) provides the guidelines for planning and managing the NCSC administration for district and school administrators. The *Directions for Test Administration* manuals by grade and form provide specific directions for TAs. The NCSC test is administered under untimed testing conditions. The current design is modeled after the intended adaptive design, with test forms built in two untimed sessions.

Evidence identification and accumulation to support the claims and validity argument were the major considerations in the development of the test blueprints. Mapping items to the claims and examining how items aggregated to provide valid inferences about students' allowed NCSC to evaluate the overall quality of the assessment system.

Conclusions and Implications

Using components of ECD requires a deep understanding of the students, standards, and evidence needed to support an accessible and technically sound assessment. Additional time in test development is needed to make explicit the targeted KSAs and the additional non-targeted KSAs that students will need to access the targeted grade-level standards. The most difficult challenge of developing the AA-AAS was creating a conceptual model that allowed for variation in the complexity of rigorous academic content. ECD uses a construct-centered approach, which required NCSC to ask what complex KSAs should be assessed based on the acquisition of academic content by SWSCD, and what is necessary for students to progress toward college, career, and community readiness. Graduated complexity models, using the tiers, provided item writers with systematic methods of varying content and supports/scaffolds to ensure that all

SWSCD have access to grade-level content standards. The nature of the construct and knowledge of the students guided the construction of the AA-AAS and development of construct-based scoring criteria.

The biggest advantage in using ECD layers was the use of common language and a common framework to facilitate communication among special educators, measurement experts, and policy makers. AA-AAS development requires experts who know the students, know the content, and know assessment development. Implementing key components of ECD created representations that did not constrain the conversations and processes, while integrating interdisciplinary conclusions into the overarching argument (Mislevy & Riconscente, 2005).

Overall, the empirical data from the field test tended to support the NCSC conceptual model for item development. As students and teachers become more familiar with the academic content and expectations, an additional examination of item tiers will determine whether there is greater separation in Rasch IP at the higher complexity levels. In the end, the answer to the essential question in this paper of “Did it work?” is, “Yes it did.”

References

Baxter, G., & Mislevy, R. J. (2005). *The case for an integrated design framework for assessing science inquiry* (PADI Technical Report 5). Menlo Park, CA: SRI International.

Browder, D., Flowers, C., Saunders, Bethune, & Spooner (2013). *Teaching students with significant cognitive disabilities the Common Core State Standards.* Unpublished manuscript.

Browder, D., & Lee, A. (2014). *A conceptual mathematics model for students with significant cognitive disabilities.* Unpublished manuscript.

Dolan, R. P., Rose, D. H., Burling, K., Harms, M., & Way, D. (April, 2007). *The universal design for computer-based testing framework: A structure for developing guidelines for constructing innovative computer-administered tests*. Paper presented at the National Council on Measurement in Education Annual Meeting, Chicago, IL.

Haertel, G. & Cameto, R. (2013, April). *Designing items of graduated complexity through systematic use of variable features and scaffolding*. Paper presented at the meeting of the American Education Research Association, San Francisco, CA.

Hess, K. (2010). *The learning progressions framework designed for use with the Common Core State Standards in Mathematics K-12.* Retrieved from http://www.nciea.org/publications/Math_LPF_KH11.pdf.

Kearns, J., Kleinert, H., Harrison, B., Sheppard-Jones, K., Hall, M., & Jones, M. (2010). *What does 'college and career ready' mean for students with significant cognitive disabilities?* Lexington: University of Kentucky.

Messick, S. (1994). The Interplay of Evidence and Consequences in the Validation of Performance Assessments. Educational Researcher, Vol. 23, No. 2, pp. 13–23.

Mislevy, R. J. (2003). Argument substance and argument structure in educational assessment. *Law, Probability and Risk, 2(4),* 237–258.

Mislevy, R., & Haertel, G. (2006). *Implications of Evidence-Centered Design for Educational Testing (Draft PADI Technical Report 17).* Menlo Park, CA: SRI International.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of assessments. Measurement: Interdisciplinary Research and Perspectives, 1, 3–62.

Mislevy, R. J., & Riconscente, M. (2005). Evidence-centered assessment design: Layers, structures, and terminology (PADI Technical Report 9). Menlo Park, CA: SRI International.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*, 3–62.

No Child Left Behind Act of 2001, 20 U.S.C. § 6319 (2008).

Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). Knowing what students know: The science and design of educational assessment. Washington, DC: National Academy Press.

Towles-Reeves, Kearns, Flowers, Hart, Kerbel, Kleinert, Quenemoen, & Thurlow (2012). *Learning characteristics project report.* Retrieved from http://ncscpartners.org/Media/Default/PDFs/Resources/LCI-Project-Report-8-21-12.pdf

Towles-Reeves, E., Kleinert, H., & Muhomba, M. (2009). Alternate assessment: Have we learned anything new? *Exceptional Children, 75*, 233-252.

U.S. Department of Education (2004). Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001. Washington, DC: Author.

U.S. Department of Education (2007). Federal Register, 72, 67, p. 74625. Retrieved from https://www2.ed.gov/legislation/FedRegister/finrule/2007-2/040907a.html

Table 1

Mathematics Prioritized CCSS, CCC, Essential Understanding, and Rationale for Prioritizing for Assessment Content

Component	Description
Claim	Student will demonstrate or explain the application of mathematical concepts
Domain	Numbers and Operations - Fractions
CCSS	3 NF Develop understanding of fractions as numbers. 3.NF.1 Understand a fraction 1/b as the quantity formed by 1 part when a whole is partitions into b equal parts; understand a fraction a/b as the quantity formed by a parts of size 1/b.
CCC	Identify the fraction that matches the representation (rectangles and circles; halves, fourths, and thirds, eighths)
Essential Understanding	Identify part and whole when item is divided – count the number of parts selected (3 of the 4 parts; have fraction present but not required to read ¾)
Rationale for Prioritizing	(1) Foundational skills that help students begin to see that things can be divided in equal shares beyond halves. (2) In Grade 3, instructional time should focus on developing understanding of fractions, especially unit fractions [fractions with numerator 1]. (3) Identify the fraction that matches the representation is a foundational skills to solving problems that involve comparing fractions by using visual fraction models. (4) Tied to CCSS Critical Areas at this grade level.

Table 2

Table of Features and Attributes of Assessment Design Patterns

Attributes	Example from 3rd Grade Mathematics
Focal KSAs	Ability to identify the fraction that matches the representation
Additional KSAs- Cognitive Background	Knowledge that (a) a fraction can be interpreted as part of a whole; (b) a fraction may be expressed as a symbol or representation; (c) the denominator of a fraction is the total number of pieces that the whole is divided into; and (d) the numerator of a fraction is the number of pieces under consideration or present in the fraction representation
Additional KSAs- Perceptual – Receptive	Ability to perceive (a) images in the stimulus material and question. (e.g., through print, objects, holistic description, Braille, audio description, tactile images) (Image in this case means a picture, drawing, table, map, graph, or photograph and not a mental image); (b) perceive physical objects required for the task (including assessment materials and response cards –e.g., see physical objects used to relate a story).This includes assessment materials and response cards; (c) the linguistic components of the stimulus material and question. (e.g., through print, objects, audio, Braille, tactile images).
Additional KSAs- Language and Symbols	Ability to (a) comprehend text, symbols, images, or objects. (Image in this case means a picture, drawing, table, map, graph, or photograph, and not a mental image); (b) decode text, symbols, tactile images, images, or objects. (Image in this case means a picture, drawing, table, map, graph, or photograph, and not a mental image); (c) recognize text, symbols, tactile images, images, or objects. (Image in this case means a picture, drawing, table, map, graph, or photograph, and not a mental image); and (d) understand English vocabulary and syntax. (If the student doesn't have the linguistic competency then it would be hard to support. If a student speaks another language then a bilingual translator can be used)
Characteristics Features	Figures will be limited to rectangles and circles; Fractions will be limited to halves, thirds, fourths, and eighths.
Variable Features: Language and Symbols	(a) Embedded support for vocabulary and symbols. (e.g., technical and non-technical glossary, hyperlinks/footnotes to definitions, illustrations, background knowledge); (b) Highlight essential elements, words, or phrases; (c) Level of abstraction required of student. (e.g., concrete objects, images, text); (d) New vs. pre-taught vocabulary and symbols; (e) Use of multiple representations. (e.g., physical models, demonstrations, acting out scenarios); and (f) Read language and symbols aloud.

Table 3

Selected Attributes across Item Tiers

Attributes	Tier 4	Tier 3	Tier 2	Tier 1
COGNITIVE BACKGROUND KNOWLEDGE				
Using both a symbolic and a visual representation of a fraction, remind student that the denominator of a fraction is the total number of pieces that the whole is divided into (numbers in the example should not be the numbers used in the prompt or selected response options)	No	No	No	No
Provide students with both a symbolic and a visual representation of the same fraction (numbers in the example should not be the numbers used in the prompt or selected response options)	No	No	Yes	No
AFFECTIVE				
Task options for engagement: enhance relevance, value, and authenticity of tasks (task refers to the assessment items, stimulus "story", and materials) In writing: create a letter to a friend, use stories with their own names or names of classmates, use content out of students' personal life	Yes	Yes	Yes	No
Teacher options for providing supports for attention and engagement: provide varied levels of challenge and support	No	No	No	No
SKILL AND FLUENCY				
Supports for manipulating digital/electronic equipment (e.g., pointers, teacher manipulation of equipment, spoken commands, stylus for input, larger keyboard/buttons, adaptive mouse)	Yes	Yes	Yes	Yes
Practice with familiar equipment	No	No	No	No

Tables 4: *Mathematics Form Summary Statistics*

Grade	Form	Total Points	Student N	Mean Raw Score	SD	Min	Max	p-value Mean	p-value SD	Alpha
3	1	40	178	19.66	7.49	1	40	.51	.12	.85
	2	40	108	21.00	7.94	5	38	.52	.14	.88
	3	40	100	18.60	7.63	0	40	.49	.14	.86
	4	40	107	20.51	7.53	0	37	.51	.13	.86
4	1	40	205	17.30	6.69	2	39	.45	.14	.82
	2	40	91	17.77	6.07	0	31	.45	.15	.78
	3	40	119	18.85	5.25	5	33	.47	.16	.69
	4	40	127	16.80	5.46	2	32	.44	.16	.72
5	1	40	189	18.17	5.09	1	35	.47	.15	.66
	2	40	78	18.03	7.14	1	40	.47	.16	.85
	3	40	95	17.60	5.08	3	31	.45	.15	.67
	4	40	101	18.06	5.58	6	36	.46	.16	.73
6	1	40	187	20.02	6.42	4	40	.50	.14	.80
	2	40	102	19.61	6.72	1	36	.50	.14	.82
	3	40	90	19.77	6.69	6	38	.50	.13	.81
	4	40	109	19.70	6.30	7	38	.49	.14	.79
7	1	40	175	18.86	5.72	6	37	.48	.16	.75
	2	40	109	20.26	7.67	0	37	.51	.15	.87
	3	40	103	20.70	6.92	1	38	.51	.16	.83
	4	40	154	19.25	6.17	8	40	.50	.15	.78
8	1	40	203	20.51	7.38	1	39	.52	.12	.85
	2	40	108	20.16	6.40	2	36	.51	.13	.79
	3	40	77	17.62	6.62	6	40	.47	.14	.81
	4	40	123	17.11	7.45	0	39	.48	.13	.85
11	1	40	210	19.25	7.15	2	39	.49	.12	.84
	2	40	102	17.51	6.35	3	34	.47	.12	.79
	3	40	60	19.13	6.04	9	37	.50	.12	.76
	4	40	124	19.33	7.97	3	39	.49	.12	.87

Table 5: *Average Rasch Item Parameter Measure by Grade and Tier*

Grade	Tier	N	Measure	SD	Grade	Tier	N	Measure	SD
3rd	1	20	-.8168	.3740	7th	1	20	-1.0437	.5024
	2	35	.0038	.5469		2	35	.0337	.5293
	3	35	.3198	.4776		3	35	.3662	.3489
	4	10	.5007	.5054		4	10	.6876	.4427
4th	1	20	-.9855	.5698	8th	1	20	-.8709	.7524
	2	35	.2204	.5664		2	35	.1488	.4697
	3	35	.2304	.4402		3	35	.1867	.4878
	4	10	.3931	.4988		4	10	.5676	.4500
5th	1	20	-1.1238	.4678	11th	1	20	-.8505	.4756
	2	35	.0302	.4008		2	35	.1380	.3697
	3	35	**.5073**	**.4007**		3	35	.2219	.3112
	4	10	**.3665**	**.4335**		4	10	.4411	.3955
6th	1	20	-.9017	.3993	All	1	100	-.9418	.5186
	2	35	.0572	.5625		2	100	.0903	.4971
	3	35	.3318	.4622		3	100	.3092	.4299
	4	10	.4419	.5957		4	100	.4855	.4688

Table 6: *Effect Sizes and Statistical Significance*

Grade	T1 v T2	T1 v T3	T1 v T4	T2 v T3	T2 v T4	T3 v T4
3rd	1.78 **	2.67 **	3.00 **	.62 *	.94 *	.37
4th	2.12 **	2.41 **	2.58 **	.02	.32	.35
5th	2.66 **	3.76 **	3.31 **	1.19 **	.81 *	.34
6th	1.99 **	2.86 **	2.7 **	.54 *	.66 *	.21
7th	2.09 **	3.31 **	3.66 **	.76 **	1.35 **	.81
8th	1.67 **	1.71 **	2.39 **	.08	.91 *	.81
11th	2.34 **	2.73 **	2.97 **	.25	.79 *	.62
All	2.03 **	2.64 **	2.89 **	.47 **	.82 **	.39 *

Note. **p<.01, *p<.05, T1=tier 1, T2 = tier 2, T3=tier 3, and T4=tier 4. Effect sizes were calculated using Cohen's *d*.

Table 7: *The Percentage of Items Performing as Expected and Average Difference in Rasch IP*

Grade		T1 v T2	T1 v T3	T1 v T4	T2 v T3	T2 v T4	T3 v T4
3rd	Pattern Met	88%	100%	100%	71%	70%	**22%**
	Logit difference	-0.9044	-1.075	-1.3044	-0.3653	-0.4642	**0.1844**
	N	16	17	5	28	10	9
4th	Pattern Met	93%	94%	100%	55%	60%	50%
	Logit difference	-1.0951	-0.9816	-1.1963	-0.0603	-0.1806	-0.1744
	N	15	17	4	29	10	10
5th	Pattern Met	100%	100%	100%	93%	75%	**43%**
	Logit difference	-1.109	-1.6792	-1.2735	-0.5191	-0.2231	**0.2062**
	N	9	11	6	14	8	7
6th	Pattern Met	93%	93%	100%	59%	67%	60%
	Logit difference	-0.8757	-1.1076	-1.0653	-0.2049	-0.4153	-0.1706
	N	14	15	4	27	6	10
7th	Pattern Met	100%	100%	100%	73%	88%	75%
	Logit difference	-1.1346	-1.3191	-1.7168	-0.3279	-0.7598	-0.3755
	N	14	14	6	22	8	8
8th	Pattern Met	88%	89%	80%	45%	75%	63%
	Logit difference	-1.0381	-1.0573	-1.7443	-0.0326	-0.6655	-0.2135
	N	17	18	5	29	8	8
11th	Pattern Met	94%	100%	100%	57%	75%	78%
	Logit difference	-1.0243	-1.0386	-1.3459	-0.1052	-0.2696	-0.1261
	N	16	17	5	23	8	9
All	Pattern Met	93%	96%	97%	62%	72%	56%
	Logit difference	-1.0204	-1.1486	-1.3989	-0.2055	-0.4187	-0.1016
	N	101	109	35	172	58	61

Table 8: *Average Difference in P-value and Random Chance for Tiers by Grade*

Grade	Tier 1	Tier 2	Tier 3	Tier 4
3	.16	.16	.10	.06
4	.14	.06	.06	.02
5	.18	.11	.01	.04
6	.19	.17	.11	.09
7	.21	.16	.09	.03
8	.16	.13	.12	.04
9	.15	.12	.10	.06
11	.17	.13	.08	.05

Table 9: *Results of Principal Factor Analyses*

Grade	Items	1st Dimension	2nd Dimension	3rd Dimension
		%	%	%
3	100	33.3	2.50	1.80
4	100	27.0	3.00	2.10
5	100	23.5	3.70	2.80
6	100	27.5	3.00	2.20
7	100	32.6	3.30	2.50
8	100	32.7	2.10	1.60
11	100	32.8	2.20	1.80

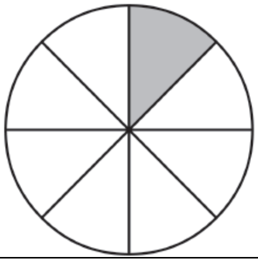
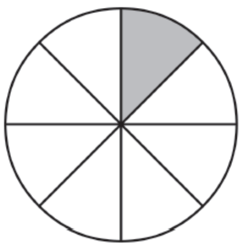
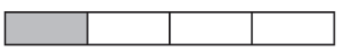
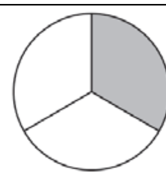
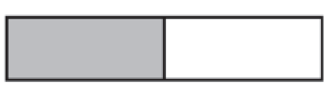
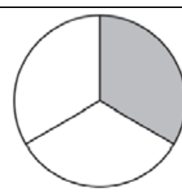
Tier 4	This item is about fractions. This fraction circle is divided into 8 equal parts. 	What part of the fraction circle is shaded? *Point to and read each option to the student.* A. $\frac{1}{2}$ B. $\frac{1}{4}$ C. $\frac{1}{8}$
Tier 3	This item is about fractions. This fraction bar is divided into 4 equal parts. *Point to each part.*  There is 1 part that is shaded. This fraction bar shows one-fourth. This fraction circle is divided into 8 equal parts. 	What part of the fraction circle is shaded? *Point to and read each option to the student.* A. $\frac{1}{2}$ B. $\frac{1}{4}$ C. $\frac{1}{8}$
Tier 2	This item is about fractions. This fraction bar is divided into 4 equal parts. *Point to each part.*  There is 1 part that is shaded. *Point to the shaded part.* This fraction shows that 1 of the 4 parts is shaded. *Point to the fraction.* $\frac{1}{4}$	 There is one part shaded. What part of the fraction circle is shaded? *Point to and read each option to the student.* A. $\frac{1}{2}$ B. $\frac{1}{3}$ C. $\frac{1}{4}$
Tier 1	This item is about fractions. This rectangle is divided into 2 equal parts. *Point to each part.*  There is 1 part that is shaded. *Point to the shaded part.* Which circle shows 1 out of 3 parts is shaded? *Point to each option.*	 A.  B.

Figure 1: Examples of items at the four levels of complexity (tiers).

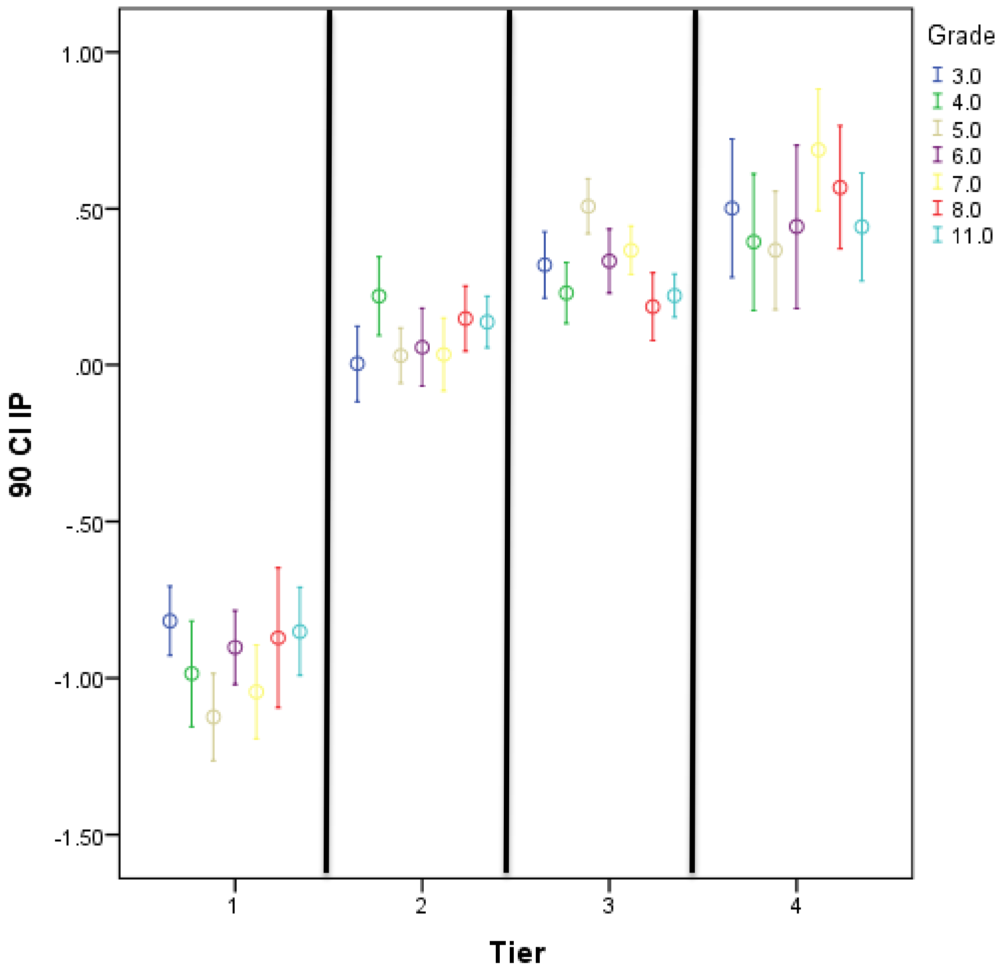


Figure 2. Confidence interval (90%) for Rasch IP measures at the different tiers.