

Comparative Alignment Analysis of Four Interim Assessment Programs with Oklahoma State ELA, Mathematics, and Science Academic Standards

July 9, 2019

Sara Christopherson
Wisconsin Center for Education Research (WCER)
School of Education
University of Wisconsin, Madison



**Wisconsin Center for
Education Research**

SCHOOL OF EDUCATION

UNIVERSITY OF WISCONSIN-MADISON

Acknowledgements

Framework Analyses:

Anna Hardway, Texas
Linda Hall, Washington D.C

English Language Arts Item Analyses:

Kymyona Burke, Mississippi
Hannah Graham, Illinois
Cindy Jacobson, Wisconsin

Mathematics Item Analyses:

Lisa Ashe, North Carolina
Lynn Raith, Pennsylvania
Diana Snider, Florida

Science Item Analyses:

Karen Mesmer, Wisconsin
Nicole Page, Georgia
Heather Toliver, Georgia

Contributors to this project have extensive experience with interim, formative, statewide summative, national, and international assessment programs. All contributors have classroom experience and/or work closely with K-12 education and are subject matter experts in their discipline. All reviewers listed above have contributed expertise to a variety of evaluative assessment-related work and some have up to 18 years of experience with alignment analysis. They bring perspectives from across the country.

Overview

In December 2018, Oklahoma Department of Education released a Request for Information (RFI) to assessment vendors of interim assessment programs commonly used within the state. Four assessment vendors responded with the requested information for ELA and mathematics: MasteryConnect submitted information for Navigate Item Bank, Measured Progress submitted information for eMPower Assessments), NWEA submitted information for MAP Growth, and Renaissance Learning submitted information for Renaissance Star 360. MasteryConnect and NWEA additionally submitted samples of science items. (Note that the Navigate Item Bank, submitted for review by MasteryConnect, is owned solely by Certica Solutions and is available through various assessment platforms, including MasteryConnect.)

The purpose of this report is to provide Oklahoma school districts with a detailed description of the results of an independent expert comparative analysis of alignment-related assessment claims for these four interim assessment programs commonly used within the state. The two-part study involved an analysis of the test program framework documentation provided by each assessment vendor and an item-level content analysis. External panelist ratings from the item-level analysis were compared with internally assigned metadata associated with 100-item samples from grade 4 and grade 7 item banks. Due to budgetary constraints, two grades were selected for this study. Grades 4 and 7 were chosen because they represent the midpoints of the OSTP grade bands of 3-5 and 6-8. The information within this report can be used by stakeholders to aid in decision-making processes.

The vendor's internal metadata specifies the vendor's intended assessment target for each item, including the standard and Depth of Knowledge (DOK) level of the item. With the assumption that the assessment programs are basing their claims about alignment on the internally assigned item metadata, the degree to which independent external reviewers agree with the internally assigned item metadata can provide some evidence to potentially substantiate these claims. Results from these studies provide information that can be used to make inferences about the capacity of a particular assessment program's item banks to yield test forms that are aligned with the state standards but does not provide direct evidence of alignment because only a stratified sample of items was analyzed for each grade and subject.

The framework analysis and item-level analysis took place remotely over the months of March - May, 2019. For each subject area, three independent external reviewers with content expertise, assessment expertise, and experience with the DOK framework reviewed the items.

Assessment vendors provided different types and scopes of information about their testing programs. All assessment vendors claimed that their items were aligned to the Oklahoma Academic Standards but did not provide specific information about how they defined “alignment” with the standards or what specific criteria were used to evaluate alignment in the test development process. The information from the framework analysis was organized into three categories for each testing program to allow for comparison of the assessment programs’ **Targets and Purposes, Alignment Claims, and Program Features**.

For each assessment vendor, item-level analysis results varied by grade and/or subject. For ELA and mathematics, external reviewer agreement with test vendor DOK assignments varied from 50% to 98%. External reviewers found a close match with the test vendor standard correlation for 59% to 100% of ELA and mathematics items. For science, reviewers found no items that directly targeted a science standard. Reviewers found that around 25% (NWEA) or around 15% (Navigate Item Bank) of the science items reviewed had some sort of relationship with the standards, but the majority of items for both assessment programs and for both grades were determined to have only a limited relationship with Oklahoma’s three-dimensional science standards.

Introduction

In Spring of 2019, an independent external analysis of interim assessment programs was conducted through the Wisconsin Center for Education Research. The analysis focused on factors related to the alignment of the interim assessments with the Oklahoma Academic Standards. Alignment of standards with assessments that purport to assess the standards is a core tenet of professional practice as recognized in Standards for Educational and Psychological Testing (AERA/APA/NCME, 2014). A high-quality assessment program involves many additional characteristics. For example, a high-quality assessment program should have a clear statement of purpose and intended use(s), demonstrate thorough and expert item and test development specifications, employ experienced item writers, include thorough editorial reviews, and implement results of bias and sensitivity analyses, along with other evaluations. This analysis was focused on alignment-related factors.

The two-part study involved an analysis of the test program framework documentation provided by each assessment vendor and an item-level content analysis. Two reviewers with extensive experience in large-scale assessment, one with expertise in English language arts (ELA) and the other in mathematics, analyzed the documentation provided by each of the four testing programs that responded to Oklahoma's RFI. The framework documentation provided was focused on ELA and mathematics programs; no framework analysis was conducted for science.

Three teams comprised of three independent external reviewers conducted the item-level analysis, one team for each subject area (ELA, mathematics, and science). Each team analyzed 100 items per subject area for each of grades 4 and 7 from each test developer. A content analysis of the item samples and the corresponding internal metadata is required to generate data that could be used as evidence to substantiate each assessment program's claims about what each item measures. These data can also be extrapolated out to evaluate claims about overall alignment of the assessment program with state standards. **Note that this item review did not compare the interim assessments with the Oklahoma School Testing Program (OSTP) assessments.**

Methodology

Two reviewers with extensive experience in large-scale assessment, one with expertise in English language arts and the other in mathematics, analyzed the documentation provided by each of the testing programs that responded to Oklahoma's RFI. These external experts gleaned technical information related to alignment claims from the assessment program documentation provided that may be useful to educators in Oklahoma for comparative purposes. Results are reported in the Framework Analysis Results section within this document.

In addition to the framework analysis, teams of independent external reviewers analyzed 100 items per subject area for each of grades 4 and 7 from each test developer. For each subject (ELA, mathematics, and science) three reviewers with content expertise, experience with DOK, and experience with alignment analyses reviewed items. The primary purpose of this item-level review was to compare the standard and DOK assignments given in the vendors' internal metadata with the judgement of independent external reviewers. The vendor's internal metadata specifies the vendor's intended assessment target for each item, including the standard and DOK level of the item. For each item, reviewers considered the standard correlation provided by the assessment vendor. Reviewers rated the standard correlation as **no match**, **plausible match**, or **match**. An item was rated as a *match* if it was possible to make a direct inference about student knowledge, skills, or abilities (KSAs) as relates to the correlated [Oklahoma Academic Standard](#). The item did not need to assess every single aspect of the correlated standard, but it did need to directly measure some core component of the standard. An item was rated as a *plausible match* if it was possible to make some sort of relevant inference about student KSAs as relates to the correlated standard, even if indirect. An item was rated as *no match* if it was not possible to make an inference about student KSAs as related to the correlated standard. Two reviewers independently reviewed each item. A third content analyst then reviewed any items for which there was a difference in coding and served as an arbiter.

Reviewers were instructed to focus primarily on evaluating the internal metadata but were able to leave qualitative comments about items as needed. Reviewers made notes about some items and flagged some items for removal from the item bank because of a Source of Challenge issue. A Source of Challenge is a circumstance in which a student might get an item correct or incorrect for the wrong reason. Results are reported in the Item-Level Analysis: Results and Discussion section within this document.

Two of the assessment vendors, Measured Progress and NWEA, have participated in the DOK Partnership Program offered by the WCEPS' WebbAlign Program. Through this partnership, assessment development staff have the opportunity to participate in annual professional development sessions as well as submit assessment items for third-party review and feedback. The feedback can be used to inform edits to items as well as adjustments to overall test programs. Extent of participation and implementation of feedback varies by partner. The goal of the DOK Partnership Program is to promote consistent, appropriate use of DOK as well as support assessment developer goals for alignment with standards. Participation in the Program indicates commitment to ongoing improvement but does not guarantee aligned assessments.

Framework Analysis Results

The information gleaned from the framework analysis was organized into three categories for each testing program: **Targets and Purposes, Alignment Claims, Program Features.**

The **Targets and Purposes** section reports the assessment targets and purposes as defined by the assessment program. For example, some programs claim to give results designed to inform instructional decision-making. Others claim a purpose of monitoring student progress, which might be linked to learning progressions, or to track student mastery of certain learning standards. Another purpose may be to predict future performance on other summative or college entrance tests (such as the OSTP, the SAT, or the ACT). All programs claim to target the Oklahoma Academic Standards. If any additional information about (or organization of) assessment targets was provided by test programs in the materials provided, this information is included in the Targets and Purposes section for that vendor.

Each of the programs makes certain claims about alignment, which may be defined slightly differently by the various programs. In this report we consider alignment to describe the relationship between the standards (assessment targets) and the interim assessments. Specifically, this means the degree to which the interim assessments address the depth and breadth of the corresponding standards. In the category of **Alignment Claims** we report the claims made by each of the assessment programs regarding their alignment with various sets of standards, and any evidence they provide that undergird those claims. In the Item-Level Analysis: Results and Discussion section, each test programs' internal coding data are provided alongside the external reviewers' findings.

The **Program Features** for each vendor includes details such as item formats, numbers of items on student tests, testing time, the size of the item bank, and whether or not the program is computer-adaptive and/or administered online.

Most of the assessment documentation provided by assessment vendors related to mathematics and ELA programs, although some programs referenced science and social studies assessments as well. The results of the framework analysis are reported for mathematics and ELA only and in alphabetical order by assessment program name: eMPower Assessments (Measured Progress), MAP Growth (NWEA), Navigate Item Bank (submitted by MasteryConnect), and STAR 360 (Renaissance Learning).

eMPower Assessments (Measured Progress)

a. Assessment Targets and Purposes

According to Measured Progress, eMPower Assessments are intended for use up to three times a year to measure student progress and growth towards grade-level expectations, plan instruction using interim score profiles, and predict performance on the OSTP. At the eighth grade level the data are intended to be used to predict performance on the SAT suite of assessments.

Table 1. eMPower Assessment Targets

Assessment	Domains assessed
Math	<p>Grade 3-5:</p> <ul style="list-style-type: none"> • Understanding and applying operations with whole numbers and fractions • Early algebraic reasoning skills • Basic measurement, geometric, and data analysis skills <p>Grades 6-8:</p> <ul style="list-style-type: none"> • Rational Numbers • Algebraic Thinking • Proportional Reasoning • Initial understanding of numeric functions
ELA	<p>Reading</p> <ul style="list-style-type: none"> • Craft and Structure • Integration of Knowledge and Ideas • Comprehension • Analysis and Interpretation • Literary Text and Informational Text <p>Writing and Language</p> <ul style="list-style-type: none"> • English Language Conventions • Narrative Writing Analysis • Expository Writing Analysis • Argument Writing Analysis

Measured Progress also claims to evaluate mathematical practices.

b. Alignment Claims

The eMPower mathematics test items "show a high degree of alignment to the OAS," according to Measured Progress. In cases in which items are not aligned, the vendor claims that the items may target more rigorous content or content that is at a different grade level. For example, per vendor description, the grade 4 Mathematics item pool may contain, in addition to items aligned to grade 4 standards in the OAS, items aligned to a higher grade's standards in the OAS or items partially aligned to the grade 4 standards in the OAS but at a higher level of rigor.

They also claim a predictive relationship between the eMPower Assessments and the PSAT and the SAT.

c. Program Features

Per vendor documentation, eMPower Assessments were built according to research from the National Governors' Association (NGA) and the Council of Chief State School Officers (CCSSO) on the knowledge and skills students must have in order to be prepared for high school and beyond in reading, writing, and mathematics. The CCSSO standards are similar to content standards in many states.

Table 2. Measured Progress Assessment Features

Assessment	Item Type(s)	Administration time	Assessment Type	Number of items
Measured Progress eMPower Math	Multiple choice, multiple select, evidence-based selected response	55-65 min	Fixed form; Online delivery	32-37
Measured Progress eMPower ELA	Multiple choice, multiple select, evidence-based selected response, extended writing prompt	Reading 50-60 min; Language Usage 35-40; Writing 60 min	Fixed form; Online delivery	17-18 reading + 23 language

Measured Progress claims to use Evidence Centered Design (ECD) to develop test and item specifications of the eMPower Assessments. ECD is a conceptual framework for the design, development, and implementation of large-scale assessments that elicit evidence to support inferences about what students know and can do.

MAP Growth (NWEA)

a. Assessment Targets and Purposes

MAP Growth assessments are intended to measure student progress in grades 3-8 toward "proficiency" on the OAS, to inform intervention and instruction, and to "promote student growth." Per NWEA, they can be used to measure growth, for instruction, as a universal screener for progress monitoring, to measure performance of programs and predict performance on summative tests.

Table 3. MAP Growth Tests Assessment Targets

Assessment	Domains assessed
<p>Math Grade 2-5</p>	<p>Number and Operations</p> <ul style="list-style-type: none"> • Count, Compare, and Represent Whole Numbers • Whole Number Operations and Problem Solving • Monetary Values and Transactions • Fractions & Decimals: Real-World & Math Situations <p>Algebraic Reasoning and Algebra</p> <ul style="list-style-type: none"> • Patterns in Real-World & Mathematical Problems • Expressions, Equations, and Inequalities <p>Geometry and Measurement</p> <ul style="list-style-type: none"> • Two- and Three-Dimensional Figures • Understand Measurable Attributes • Solve Problems Involving Time <p>Data and Probability</p> <ul style="list-style-type: none"> • Data Analysis
<p>Math Grade 6-12</p>	<p>Number and Operations</p> <ul style="list-style-type: none"> • Real and Complex Numbers • Operations and Problem Solving with Rational Numbers <p>Algebraic Reasoning and Algebra</p> <ul style="list-style-type: none"> • Expressions, Equations, and Inequalities • Functions <p>Geometry and Measurement</p> <ul style="list-style-type: none"> • Measurement, Relationships, and Coordinate Geometry • Congruence, Similarity, Right Triangles, and Logic <p>Data and Probability</p> <ul style="list-style-type: none"> • Data Analysis • Probability and Statistical Thinking

Table 3 Cont'd. MAP Growth Tests Assessment Targets

Assessment	Domains assessed
<p>Reading, Grades 2-12</p>	<p>Reading Processes: Read and Comprehend Texts</p> <ul style="list-style-type: none"> • Main Ideas and Supporting Details; Text Features • Genre <p>Critical Reading: Interpret and Evaluate Texts</p> <ul style="list-style-type: none"> • Author's Perspective, Purpose, and Point of View • Inferences and Conclusions; Text Structures • Literary Elements and Devices <p>Vocabulary</p> <ul style="list-style-type: none"> • Word Relationships; Word Parts • Context Clues; Academic Vocabulary; Reference Materials
<p>Language Usage, Grades 2-12</p>	<p>Writing Process</p> <ul style="list-style-type: none"> • Research • Prewriting, Drafting • Revising, Editing • Spelling <p>Writing Modes</p> <ul style="list-style-type: none"> • Narrative • Informative • Opinion/Argumentative <p>Language</p> <ul style="list-style-type: none"> • Capitalization, Punctuation • Grammar, Usage • Sentence Structure and Variety

b. Alignment Claims

Per NWEA documentation, NWEA researchers conduct regular linking studies to examine the correspondence between MAP Growth assessments and statewide summative assessments used to measure student achievement. The NWEA Psychometric Solutions team recently completed a linking study to connect MAP Growth RIT scores to the OSTP General Assessments.

NWEA has conducted third-party alignment evaluation of the MAP Growth assessment with the Oklahoma Academic Standards and has made results available. NWEA is involved in studies linking MAP Growth to the ACT and the SAT at grades 5-9, based on which the vendor is prepared to make predictive claims.

c. Program Features

MAP Growth assessments are computer adaptive. The item bank contains more than 38,000 items, and the entire bank is available within a content area for a student's assessment.

Table 4. NWEA MAP Growth Assessment Features

Assessment	Item Type(s)	Administration time	Assessment Type	Number of items
NWEA Math	Multiple choice, multiple select, evidence-based selected response, and technology enhanced items	Untimed; 40-60 minutes average	Computer adaptive	variable
NWEA ELA	Multiple choice, multiple select, evidence-based selected response, and technology enhanced items	Untimed; 40-60 minutes average	Computer adaptive	variable

Navigate Item Bank (submitted by MasteryConnect)

a. Assessment Targets and Purposes

Items from the Navigate Item Bank were submitted by MasteryConnect. The Navigate Item Bank is owned solely by Certica Solutions and is available through a number of assessment platforms across the country, including MasteryConnect. The Navigate Item Bank is one of multiple content solutions and partnerships with vendors that are maintained by MasteryConnect.

MasteryConnect defines the Navigate Item Bank assessment system's primary goal as tracking student mastery of learning standards. Mathematics and ELA items are available for Kindergarten through 12th grade. MasteryConnect claims that the Navigate Item Bank includes items that target all of the Oklahoma Academic Standards. No domains or other categories of assessment targets are identified outside of the state standards.

According to information provided by MasteryConnect, the assessment tool is intended to deliver standards-based data for educators immediately. The direct application can include "quick" classroom formative assessments, district-wide formative assessments, or benchmark tests. The resulting data is intended to inform instruction before end-of-year testing, and to track progress towards student mastery of standards.

b. Alignment Claims

MasteryConnect claims 27,846 mathematics items and 35,071 ELA items in the Navigate Item Bank are aligned to Oklahoma Academic Standards.

c. Program Features

Both customized and pre-created assessments are available. MasteryConnect claims the Navigate Item Bank contains 77,589 items that assess the Oklahoma Academic Standards, divided between mathematics, ELA, science, and social studies. They report between 1,000 and around 3,800 items in an item bank for each grade for ELA grades K-12 and between around 1,000 and over 4,000 items in an item bank for each grade for mathematics grades K-Algebra 2.

Table 5. Navigate Item Bank Assessment Features

Assessment	Item Type(s)	Administration time	Assessment Type	Number of items
Navigate Item Bank Math	multiple choice, constructed response, multipart, technology enhanced	variable	Fixed form; Online or printable	variable
Navigate Item Bank ELA	multiple choice, constructed response, multipart, technology enhanced	variable	Fixed form; Online or printable	variable

STAR 360: Early Reading, Literacy, and Math (Renaissance Learning)

a. Assessment Targets and Purposes

Renaissance Learning states that the STAR 360 assessments are intended for multiple purposes including screening to identify students in need of further instruction and to inform instructional decisions, prediction of performance on the Oklahoma School Testing Program, measurement of student growth, and measurement of proficiency on the Oklahoma state standards.

Table 6. STAR 360 Assessment Targets

Assessment	Domains assessed
STAR Early Literacy	<ul style="list-style-type: none"> • Word knowledge and skills • Comprehension strategies and constructing meaning • Numbers and operation
STAR Reading	<ul style="list-style-type: none"> • Word knowledge and skills • Comprehension strategies and constructing meaning • Analyzing literary text • Understanding author’s craft • Analyzing argument and evaluating text
STAR Math	<ul style="list-style-type: none"> • Numbers and operations • Algebra • Geometry and measurement • Data analysis, probability, and statistics

b. Alignment Claims

Renaissance claims that the STAR 360 assessment system is fully aligned with the Oklahoma Academic Standards and linked to learning progressions developed in the context of the Oklahoma standards. The STAR 360 assessments for grades 3-8 have been statistically linked to PARCC and SBAC, state summative assessments, the ACT, and the SAT. The vendor provides evidence of concurrent and predictive correlations, as well as results from several predictive studies.

c. Program Features

The STAR 360 assessment system includes computer adaptive tests that are normed across millions of users, based on five norming conventions: geographically specific norming by 4 regions within the US, school size norming by three specific ranges, socioeconomic norms in four categories, and type of school/learner.

Table 7. Renaissance Learning Assessment Features

Assessment	Item Type(s)	Administration time	Assessment Type	Number of items
STAR Early Literacy	Multiple choice, multiple select, evidence-based selected response	9 minutes average	Computer adaptive	27
STAR Reading	Multiple choice, multiple select, evidence-based selected response	Less than 20 minutes average	Computer adaptive	34
STAR Math Progress Monitoring	Multiple choice, multiple select, evidence-based selected response	Less than 14 minutes average	Computer adaptive	24
STAR Math	Multiple choice, multiple select, evidence-based selected response	Less than 25 minutes average	Computer adaptive	34

Item-Level Analysis: Results and Discussion

Item-level analysis results varied by grade and/or by subject. Overall results for mathematics are summarized in Tables 8-9 and for ELA in Tables 10-11. These results depict reviewer agreement with the test vendors' internal correlation of items to standards and internal assignment of DOK. Tables 8 (math) and 10 (ELA) show the percent of items that reviewers rated as no match, plausible match, or match with the assessment vendor's internally assigned standard. There may be several reasons why items are mismatched with standards. If items originally written to target one set of standards are repurposed for assessment of another set of standards, mismatches are common. Other causes include differences in interpretation of the intent of standards and/or items, differences in the protocols or guidelines used to determine if an item targets or assesses a standard, and human error.

Reviewers had least agreement with the standard correlations for eMPower mathematics items (G4: 59%; G7: 85% considered match; See Table 8) and least agreement with the Navigate Item Bank ELA items (G4: 67%; G7: 64% considered match; See Table 10). Recall that an item was rated as a *match* if it was possible to make a direct inference about student knowledge, skills, or abilities (KSAs) as relates to the correlated standard.

Reviewers had least agreement with the Navigate Item Bank DOK assignments (<60% for both grades and both disciplines; See Tables 9 and 11). The DOK codings for this test program were not consistent with the subject area definitions for Webb’s Depth of Knowledge (DOK). However, the distribution of items by DOK (as determined by experienced external reviewers) was similar to the other programs reviewed. The distribution of items by DOK varied slightly by program and sometimes by grade level within the same program. Tables 9 and 11 show the percent of items reviewed that external reviewers determined to require DOK 1, 2, or 3 type processes. Full definitions of DOK for each subject area are available as an appendix to this report.

Only the STAR 360 item sample included DOK 3 tasks for mathematics. Note, however, that only a small proportion of the K-12 mathematics standards expect DOK 3 work. For ELA, there is a much larger proportion of DOK 3 expectations, and therefore appropriate that the ELA item samples contained relatively more DOK 3 items/tasks than did the math item samples. All test samples contained at least some DOK 3 items in the ELA sample. The percentage of DOK 3 items in grade 4 samples ranged from 2% to 15% and in grade 7 samples ranged from 7% to 22%. The fewest DOK 3 items were found in the grade 4 MAP Growth sample (2%) and the most in the grade 7 MAP Growth sample (22%). When interpreting results, it is important to remember that a relatively small proportion of the item banks for each assessment program was reviewed.

Table 8. Math Interim Assessments: Agreement with Vendor Standard Assignments

Math	% Items with NO MATCH with given standard	% Items with PLAUSIBLE MATCH with given standard	% Items with MATCH with given standard
eMPower G4	18%	25%	59%
eMPower G7	9%	6%	85%
MAP Growth G4	2%	0%	98%
MAP Growth G7	4%	6%	90%
Navigate G4	2%	1%	97%
Navigate G7	1%	2%	97%
STAR 360 G4	6%	12%	82%
STAR 360 G7	1%	0%	100%

Table 9. Math Interim Assessments: Agreement with Vendor DOK Assignments and DOK Distribution

Math	DOK % agreement with internal coding	% items DOK 1 (based on external review)	% items DOK 2 (based on external review)	% items DOK 3 (based on external review)	% items DOK 4 (based on external review)
eMPower G4	86%	50%	50%	0%	0%
eMPower G7	95%	24%	76%	0%	0%
MAP Growth G4	88%	52%	48%	0%	0%
MAP Growth G7	93%	31%	69%	0%	0%
Navigate G4	57%	31%	69%	0%	0%
Navigate G7	59%	23%	77%	0%	0%
STAR 360 G4	92%	4%	77%	13%	6%
STAR 360 G7	61%	38%	49%	13%	0%

Table 10. ELA Interim Assessments: Agreement with Vendor Standard Assignments

ELA	% Items with NO MATCH with given standard	% Items with PLAUSIBLE MATCH with given standard	% Items with MATCH with given standard
eMPower G4	3%	14%	83%
eMPower G7	1%	9%	90%
MAP Growth G4	0%	17%	83%
MAP Growth G7	0%	16%	84%
Navigate G4	11%	22%	67%
Navigate G7	14%	22%	64%
STAR 360 G4	5%	16%	79%
STAR 360 G7	0%	16%	84%

Table 11. ELA Interim Assessments: Agreement with Vendor DOK Assignments and DOK Distribution

ELA	DOK % agreement with internal coding	% items DOK 1 (based on external review)	% items DOK 2 (based on external review)	% items DOK 3 (based on external review)	% items DOK 4 (based on external review)
eMPower G4	98%	18%	77%	5%	0%
eMPower G7	92%	12%	81%	7%	0%
MAP Growth G4	87%	36%	62%	2%	0%
MAP Growth G7	88%	19%	59%	22%	0%
Navigate G4	50%	20%	64%	15%	0%
Navigate G7	54%	23%	55%	21%	0%
STAR 360 G4	95%	13%	74%	13%	0%
STAR 360 G7	95%	11%	75%	13%	1%

Vendors' Metadata: Agreement with External Analysis

Figures 1-4 show the distribution of items by grade and subject for each standard according to the test vendors' internal coding. The last set of bars (on the far right side of each graph) gives the percent of items for which external reviewers agreed with the internal codings. Across test forms, independent panelist agreement with internal standard correlations ranged from 57% (Navigate Item Bank, grade 4 math) to 100% (STAR 360, grade 7 math). When considering the breadth of standards shown by each test vendor's internal codings, it is important to check the extent to which the internal codings were consistent with independent codings. For example, when looking at the Navigate Item Bank's internal standard correlations for grade 4 math in Figure 1, remember that panelists disagreed with nearly half of those correlations. Thus, the distribution of items across domains that is shown likely does not reflect the actual distribution.

In general, the external reviewers found items corresponding to standards across the same domains as claimed by each assessment vendor. However, external reviewers did not always agree with the precise internal codings. For example, in Figure 1 on the following page, reviewers found items that corresponded to the same five domains as in the STAR 360 metadata, but agreed with the specific standards identified for only 79% of the items.

Figure 1. Grade 4 Math Standards with Corresponding Assessment Item(s)

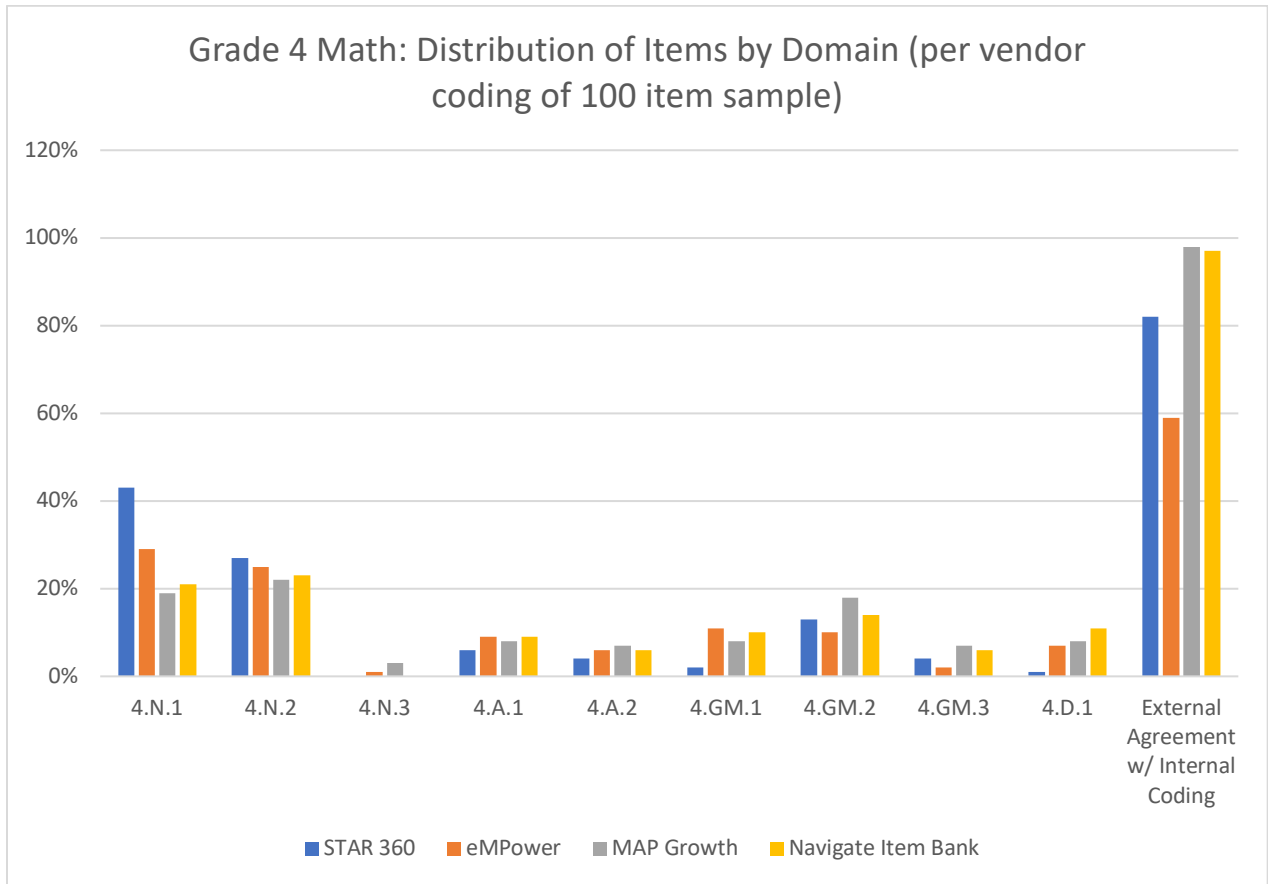


Figure 2. Grade 7 Math Standards with Corresponding Assessment Item(s)

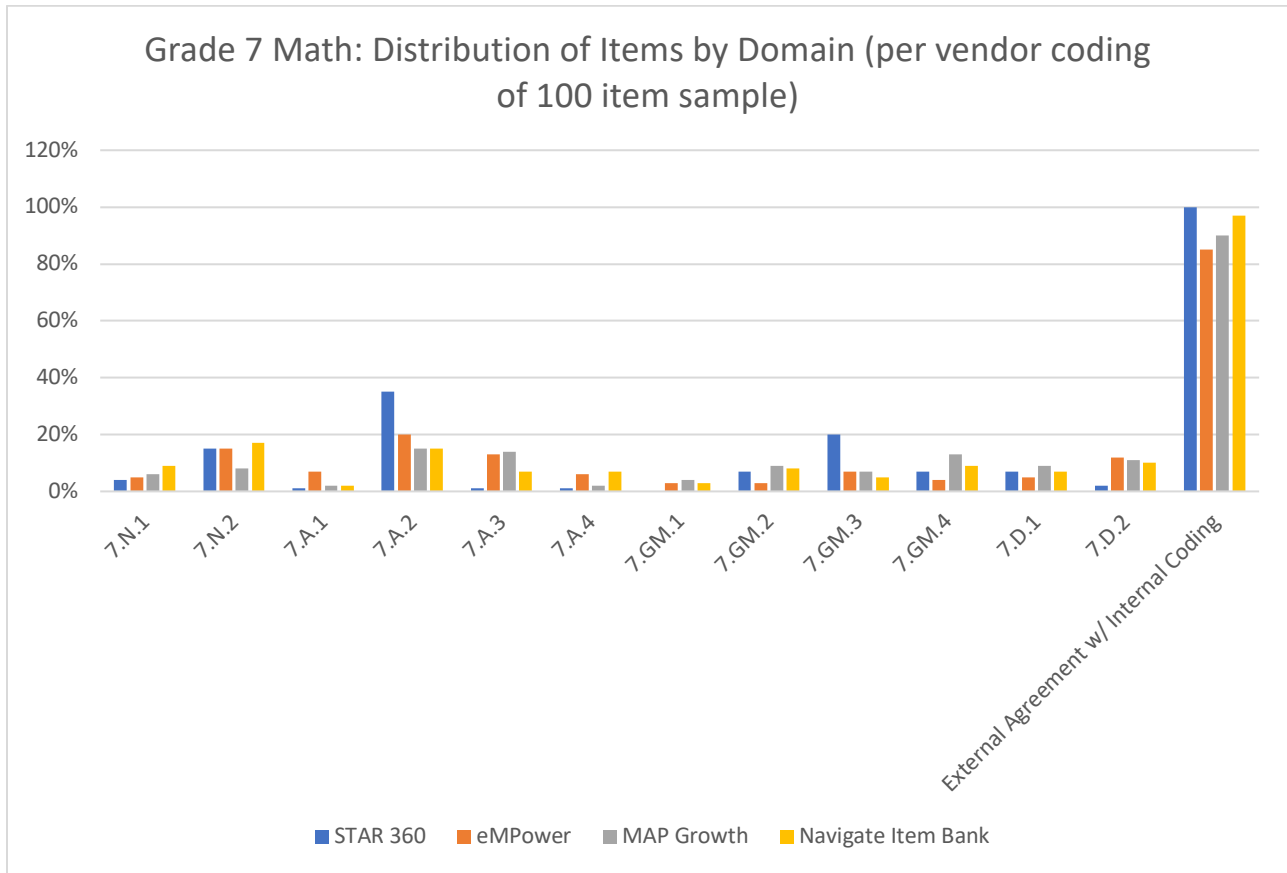


Figure 3. Grade 4 ELA Domains with Corresponding Assessment Item(s)

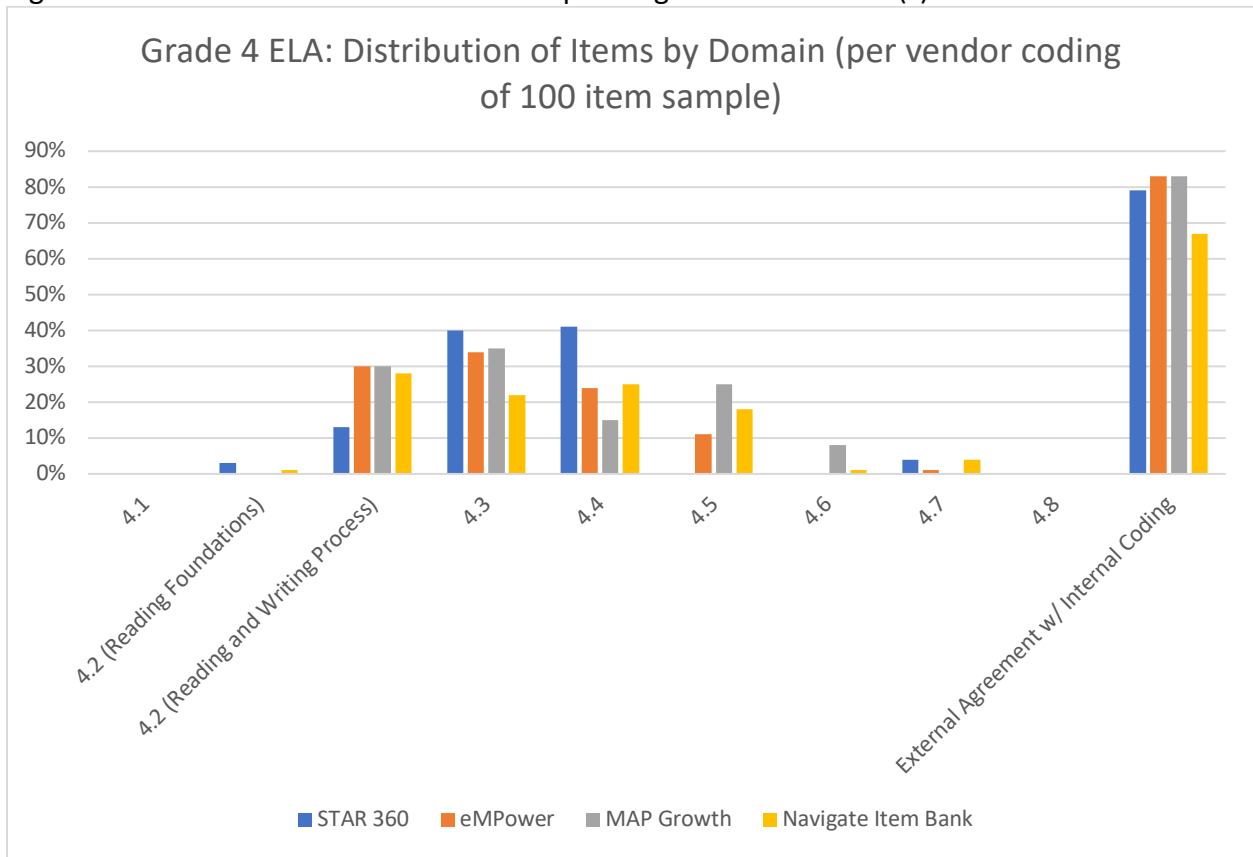
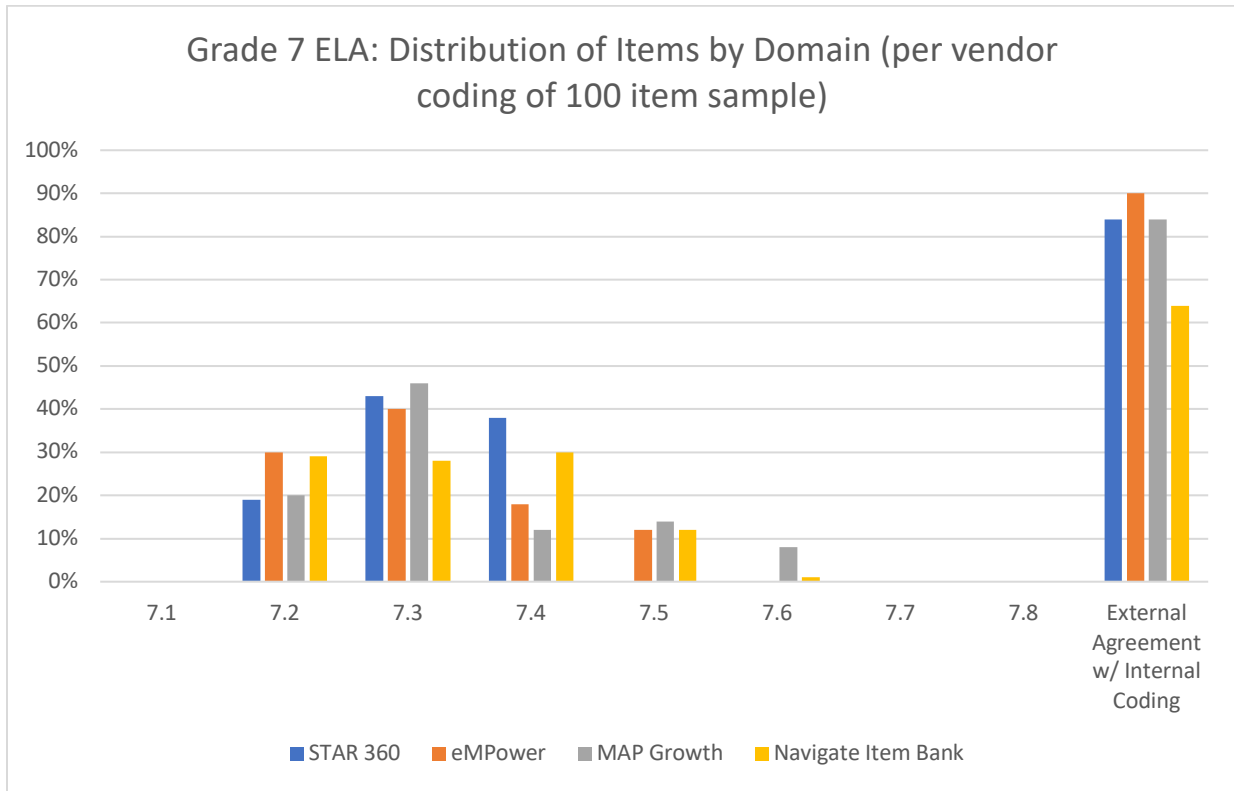


Figure 4. Grade 7 ELA Domains with Corresponding Assessment Item(s)



Tables 12-15 list the domains and standards that were unassessed by any items within each 100-item sample. The last row in each table identifies any emphasis in the set of items. For example, 31% of the Grade 4 ELA STAR 360 items targeted Standard 4.4.R.1, related to academic vocabulary. Because the sets of 100 items were stratified random samples from a larger item bank, these results should represent the larger bank. However, some vendors' item banks contain tens of thousands of items, and a 100-item sample may not adequately represent the full bank.

Table 12. Grade 4 mathematics unassessed objectives per review of 100 item samples

GRADE 4 math objectives with no corresponding item (per 100 item sample)				
Grade 4 Math	eMPower	Navigate	MAP Growth	STAR 360
4.N.1	At least one item targeted each objective	4.N.1.4	At least one item targeted each objective	At least one item targeted each objective
4.N.2	At least one item targeted each objective	4.N.2.8	4.N.2.8	4.N.2.8
4.N.3	One item targeted this objective	4.N.3.1	At least one item targeted this objective	4.N.3.1
4.A.1	At least one item targeted each objective	At least one item targeted each objective	At least one item targeted each objective	4.A.1.3
4.A.2	At least one item targeted each objective	At least one item targeted each objective	At least one item targeted each objective	At least one item targeted each objective
4.GM.1	4.GM.1.3	At least one item targeted each objective	4.GM.1.3	4.GM.1.2, 4.GM.1.3
4.GM.2	4.GM.2.3	At least one item targeted each objective	At least one item targeted each objective	At least one item targeted each objective
4.GM.3	At least one item targeted each objective	At least one item targeted each objective	At least one item targeted each objective	4.GM.3.1
4.D.1	4.D.1.2	At least one item targeted each objective	At least one item targeted each objective	4.D.1.1, 4.D.1.2
Additional Observation: EMPHASIS (if any):	no objective targeted by more than 10% of items	no objective targeted by more than 10% of items	no objective targeted by more than 10% of items	11% of items targeted 4.N.1.1; 13% of items targeted 4.N.2.2

Table 13. Grade 7 mathematics unassessed objectives per review of 100 item samples

GRADE 7 math objectives with no corresponding item (per 100 item sample)				
Grade 7 Math	eMPower	Navigate	MAP Growth	STAR 360
7.N.1	7.N.1.2	At least one item targeted each objective	At least one item targeted each objective	7.N.1.1
7.N.2	7.N.2.1	7.N.2.1	7.N.2.1, 7.N.2.2, 7.N.2.1, 7.N.2.5	7.N.2.1, 7.N.2.2
7.A.1	At least one item targeted each objective	At least one item targeted each objective	7.A.1.1	7.A.1.2
7.A.2	At least one item targeted each objective	7.A.2.4	7.A.2.4	7.A.2.2
7.A.3	At least one item targeted each objective	At least one item targeted each objective	At least one item targeted each objective	7.A.3.1, 7.A.3.2
7.A.4	7.A.4.2	At least one item targeted each objective	7.A.4.2	7.A.4.2
7.GM.1	7.GM.1.2	7.GM.1.1	7.GM.1.1	No items targeted these objectives
7.GM.2	At least one item targeted each objective	At least one item targeted each objective	At least one item targeted each objective	7.GM.2.1
7.GM.3	At least one item targeted each objective	At least one item targeted each objective	At least one item targeted each objective	7.GM.3.1
7.GM.4	7.GM.4.3	At least one item targeted each objective	At least one item targeted each objective	7.GM.4.1, 7.GM.4.3
7.D.1	At least one item targeted each objective	At least one item targeted each objective	At least one item targeted each objective	At least one item targeted each objective
7.D.2	At least one item targeted each objective	At least one item targeted each objective	At least one item targeted each objective	7.D.2.1, 7.D.2.3
Additional Observation: EMPHASIS (if any):	no objective targeted by more than 10% of items	no objective targeted by more than 10% of items	no objective targeted by more than 10% of items	11% of items targeted 4.N.1.1; 13% of items targeted 4.N.2.2

Table 14. Grade 4 ELA unassessed objectives per review of 100 item samples

GRADE 4 ELA objectives with no corresponding item (per 100 item sample)				
Grade 4 ELA	eMPower	Navigate	MAP Growth	STAR 360
STANDARD 1: Speaking and Listening	No items targeted standard 1	No items targeted standard 1	No items targeted standard 1	No items targeted standard 1
STANDARD 2: Reading Foundations	No items targeted objective 2 RF	No items targeted objective 2 RF	4.2.PC, 4.2.F.1, 4.2.F.2	4.2.PC, 4.2.F.1, 4.2.F.2
STANDARD 2: Reading and Writing Process	4.2.W.4	At least one item targeted each objective	4.2.R.4	4.2.R.4, 4.2.W.1, 4.2.W.2, 4.2.W.3, 4.2.W.4
STANDARD 3: Critical Reading and Writing	4.3.R.4, 4.3.R.5, 4.3.W.3	At least one item targeted each objective	4.3.R.2	4.3.W.1, 4.3.W.2, 4.3.W.3
STANDARD 4: Vocabulary	4.4.R.2	At least one item targeted each objective	4.4.R.5, 4.4.W.1	4.4.R.2, 4.4.R.5, 4.4.W.1, 4.4.W.2
STANDARD 5: Language	4.5.R.1, 4.5.R.3, 4.5.R.5, 4.5.W.4	At least one item targeted each objective	4.5.R.3, 4.5.R.4, 4.5.R.5, 4.5.W.1 4.5.W.4	No items targeted standard 5
STANDARD 6: Research	No items targeted standard 6	4.6.R.1, 4.6.W.2, 4.6.W.3	4.6.R.1, 4.6.R.3 4.6.W.1, 4.6.W.2, 4.6.W.3	No items targeted standard 6
STANDARD 7: Multimodal Literacies	No items targeted standard 7	No items targeted standard 7	4.7.R.2, 4.7.W.1, 4.7.W.2	4.7.R.2, 4.7.W.1, 4.7.W.2
STANDARD 8: Independent Reading and Writing	No items targeted standard 8	No items targeted standard 8	No items targeted standard 8	No items targeted standard 8
EMPHASIS (if any):	13% of items targeted 4.2.W.2	17% of items targeted 4.3.R.3	no objective targeted by more than 10% of items	31% of items targeted 4.4.R.1

Table 15. Grade 7 ELA unassessed objectives per review of 100 item samples

GRADE 7 ELA objectives with no corresponding item (per 100 item sample)				
Grade 7 ELA	eMPower	Navigate	MAP Growth	STAR 360
STANDARD 1: Speaking and Listening	No items targeted standard 1	No items targeted standard 1	No items targeted standard 1	No items targeted standard 1
STANDARD 2: Reading and Writing Process	7.2.W.2	At least one item targeted each objective	7.2.W.5	7.2.W.1, 7.2.W.2, 7.2.W.3, 7.2.W.4 7.2.W.5
STANDARD 3: Critical Reading and Writing	7.3.R.1, 7.3.W.1	7.3.R.1	7.3.R.1, 7.3.R.5 7.3.W.4	7.3.R.1, 7.3.R.7 7.3.W.1, 7.3.W.2, 7.3.W.3, 7.3.W.4
STANDARD 4: Vocabulary	At least one item targeted each objective	At least one item targeted each objective	At least one item targeted each objective	7.4.W.1, 7.4.W.2
STANDARD 5: Language	7.5.R.1, 7.5.R.2, 7.5.R.3	7.5.W.3	7.5.R.1, 7.5.R.2, 7.5.R.3, 7.5.W.3	No items targeted standard 5
STANDARD 6: Research	No items targeted standard 6	7.6.R.1, 7.6.W.1, 7.6.W.4	7.6.R.1, 7.6.R.2, 7.6.W.1, 7.6.W.2, 7.6.W.3, 7.6.W.4	No items targeted standard 6
STANDARD 7: Multimodal Literacies	No items targeted standard 7	No items targeted standard 7	No items targeted standard 7	No items targeted standard 7
STANDARD 8: Independent Reading and Writing	No items targeted standard 8	No items targeted standard 8	No items targeted standard 8	No items targeted standard 8
EMPHASIS (if any):	13% of items targeted 7.3.R.6	13% of items targeted 7.3.R.3 and 14% of items targeted 7.3.R.4	12% of items targeted 7.2.R.1	22% of items targeted 7.4.R.1

Science Items

Two vendors, NWEA and MasteryConnect, submitted sets of science items for inclusion in this analysis. MasteryConnect submitted science items from the Navigate Item Bank, which is owned solely by Certica Solutions and is available through various assessment platforms, including MasteryConnect. The item analysis was focused on the relationship between the item and the internally correlated metadata. Reviewers evaluated the extent to which an item measured the knowledge, skills, and abilities detailed in the corresponding three-dimensional performance expectation (standard) identified by the vendor.

Reviewers found some sort of relationship with the Oklahoma science standards for around 25% of the science items submitted by NWEA and for around 15% of the Navigate Item Bank science items reviewed. The majority of items for both assessment programs and for both grades were determined to have no relationship with the expectations of the three-dimensional standards beyond limited overlap with the general science topics.

In general, the science items assessed factual and definitional knowledge related to science as well as conceptual understanding of science concepts and relationships. The sample from Navigate Item Bank had a greater proportion of items targeting recalled facts than did the sample from NWEA; NWEA had a greater proportion of items targeting conceptual understanding. However, neither set of items fully or cohesively targeted Oklahoma's three-dimensional science standards. Reviewers saw limited evidence of multidimensional assessment items/tasks. Reviewers also commented that even when the item targeted a science fact or topic, it was not necessarily the specific disciplinary core idea (DCI) that was included in the correlated performance expectation but instead just a science topic related to the DCI.

The Navigate Item Bank science items for each grade included items that were +/- 2 grade levels. For example, items within the grade 7 item pool could include both grade 5 items and high school items. Reviewers commented on these off-grade items, noting that the topics were often not included in the on-grade standards. NWEA items were labeled as on-grade, but reviewers commented on a number of items that they noticed assessed material related to off-grade content.

Summary

In December 2018, Oklahoma Department of Education released a Request for Information (RFI) to assessment vendors of interim assessment programs commonly used within the state. Four assessment vendors responded with the requested information for ELA and mathematics: MasteryConnect submitted information for Navigate Item Bank, Measured Progress submitted information for eMPower Assessments), NWEA submitted information for MAP Growth, and Renaissance Learning submitted information for Renaissance Star 360. MasteryConnect and NWEA additionally submitted samples of science items. (Note that the Navigate Item Bank, submitted for review by MasteryConnect, is owned solely by Certica Solutions and is available through various assessment platforms, including MasteryConnect.)

A two-part study was conducted on the materials submitted: an analysis of the test program framework documentation provided by each assessment vendor and an item-level content analysis. The framework analysis and item-level analysis took place remotely over the months of March - May, 2019. Two reviewers with extensive experience in large-scale assessment, one with expertise in English language arts (ELA) and the other in mathematics, analyzed the documentation provided by each of the four testing programs that responded to Oklahoma's RFI. The framework documentation provided was focused on ELA and mathematics programs; no framework analysis was conducted for science.

For each subject area, a team of three independent external reviewers with content expertise, assessment expertise, and experience with the DOK framework reviewed the items. External panelist ratings from the item-level analysis were compared with internally assigned metadata associated with 100-item samples from grade 4 and grade 7 item banks. **Note that this item review did not compare the interim assessments with the Oklahoma School Testing Program (OSTP) assessments.**

The goal of the work was to provide Oklahoma school districts with information about interim assessment programs commonly used within the state based on the results of an independent expert comparative analysis of alignment-related assessment claims of the test programs. Item-level analysis results varied by grade and/or subject. For ELA and mathematics, external reviewer agreement with test vendor DOK assignments varied from 50% to 98%. External reviewers found a close match with the test vendor standard correlation for 59% to 100% of ELA and mathematics items. For science, reviewers found some sort of relationship with the Oklahoma science standards for around 25% of the science items submitted by NWEA and for around 15% of the Navigate Item Bank science items reviewed. The majority of items for both assessment programs and for both grades were determined to have no relationship with the expectations of the three-dimensional standards beyond limited overlap with the general science topics.

Based on the item samples reviewed in this analysis, the assessment programs vary by grade, breadth of standards assessed, complexity of items, and other factors. Stakeholders can use the comparative information within this report as one of many pieces of information that can contribute to the choice of interim assessment program(s) that are the best fit for the needs of their district.

The WebbAlign program can be reached at webbalign@wceps.org.

References

American Educational Research Association., American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*.

Webb, N. L. (1997). Criteria for alignment of expectations and assessments in mathematics and science education. Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6. Madison, WI: University of Wisconsin.

Webb, N. L. (1999). Alignment of science and mathematics standards and assessments in four states. Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 18. Madison, WI: University of Wisconsin.

Webb, N. L. (2003). Alignment study in language arts, mathematics, science, and social studies of state standards and assessments in four states. Washington, D. C.: Council of Chief State School officers.