Mathway

A Formative Assessment Tool for Math Corps

Technical Report

September 2023

NATIONAL Science & Service COLLABORATIVE

Powerful Insights. AmeriCorps Ingenuity.

AmeriCorps

ServeMN

About the National Science & Service Collaborative

We believe partnerships between researchers, AmeriCorps programs, and communities can transform research and practice, leading to sustainable, community-driven solutions. We value a broad and inclusive definition of "collaboration" because improving societal outcomes is maximized when the tools of science, expertise of communities, and resources of AmeriCorps are deployed in a truly collaborative way.

The Center's portfolio includes projects to evaluate the impact of AmeriCorps programming, projects to advance the existing knowledge base in education, and development projects to bring new and innovative programming to communities across the nation.
https://nssc.serveminnesota.org/

About the Authors

The Mathway technical document encompasses the efforts of a variety of stakeholders, including teachers, program staff, researchers, AmeriCorps tutors, and the NSSC. The construction of the report was led by Dr. Garret Hall (Florida State University) and Dr. Peter Nelson (ServeMinnesota).

Dr. Garret Hall served as a technical lead on Mathway and is an assistant professor in the MS/EdS School Psychology program as well as the Combined Counseling and School Psychology PhD program at Florida State University. He graduated with his Ph.D. and M.S in Educational Psychology from the University of Wisconsin-Madison and received a B.A. with majors in English and Psychology from Northern Illinois University. Dr. Hall's research focuses primarily on development of academic skills (especially mathematics). This includes examining predictors of academic development (such as language and linguistic diversity as well executive functions) as well as equitable methods to assess, prevent, and intervene on academic difficulties. Dr. Hall has deep expertise in quantitative methodology (including causal inference, measurement, and longitudinal methods) and how these methods can inform school-based prevention and intervention within multi-tiered systems of supports (MTSS). Last, he is interested in how "ecological" factors within MTSS promote students' academic, social, emotional, and behavioral functioning, such as implementation fidelity and family-school involvement.

Dr. Peter Nelson is the Vice President of Impact and Innovation at ServeMinnesota where he engages with ServeMinnesota pilot programs and existing programs to ensure key principles of effective implementation and evidence-building occur. Dr. Nelson has more than 10 years of experience in research and evaluation methodology with a particular focus on the educational context. Before joining ServeMinnesota, Dr. Nelson was an Assistant Professor of School Psychology at Penn State University and over the course of his career has led the development of three math assessments and one classroom environment assessment tool. Dr. Nelson has published over 40 peer-reviewed articles in academic journals and helped secure over 2 million dollars in funding to facilitate new program development and research related to data-based decision making and academic intervention in the school setting. Dr. Nelson holds a B.A. in Psychology from St. Olaf College, an M.A. in Education from the University of Mississippi, and an M.Ed and Ph.D. from the University of Minnesota in School Psychology.

Table of Contents

Introduction to the Mathway Assessment

The Mathway technical manual provides an overview of Math Corps and technical documentation for Mathway—the primary assessment used by tutors within the Math Corps intervention model. Mathway was collaboratively developed with teachers, Math Corps content leads, researchers, and other key stakeholders. The assessment spans across grades K-8 and is administered in an online environment wherein data are collected, managed, and displayed for program use.

Mathway is explicitly designed to complement the Math Corps program by providing tutors with assessment data that are directly linked to the intervention content. Mathway is used by tutors across the country to inform student selection and exit decisions, guide service delivery, and evaluate impact on student learning.

Math Corps Theory of Change and Overview of Data Use

Mathway was developed explicitly for use within the Math Corps intervention program model. It is therefore important to understand the nature and scope of the Math Corps program as context for Mathway and its application in schools.

Math Corps was originally developed to address long-standing capacity gaps within schools in regard to the provision of evidence-based supplemental support for students with a demonstrated need for that support. The theory of change for the program is consistent across grades and is rooted in the notion that AmeriCorps members with variable backgrounds in education can be recruited and trained to provide high-dosage tutoring in schools, functionally operating as part-time or full-time members of the school community. Since its initial implementation, Math Corps has grown to serve thousands of students each year across nine states.

The growth of Math Corps is directly related to the acute need in schools for tutoring support as well as the program's long-term approach to evidence building. Math Corps has completed two separate randomized controlled trials—both of which were led by external parties—in which students served by the program performed significantly better on broad-based measures of math achievement and fact fluency assessments (Codding et al., 2022). Parker et al., 2019) In addition, Math Corps research projects have advanced understanding of important facets of high-dosage tutoring programs (Nelson et al., 2018; Nelson et al, 2019; Van Norman et al., 2017).

Math Corps Content Focus

Given the scope of Math Corps implementation, there is a great deal of variance across school partners in regard to curricula, ideological alignment, and resources. Math Corps has a "curriculum-independent" approach to content. In other words, by design, Math Corps does not directly align with the curricular distribution and sequence in schools. The rationale for this approach to program design is two-fold.

First, students in need of supplemental academic support in math—particularly in late elementary and middle school—exhibit deficits that originate with foundational content introduced in previous grade levels (Spaull & Kotze, 2015). For example, research indicates that when presented with two fractions, middle school students are unable to correctly identify which is larger better than chance probability (Siegler & Lortie-Forgues, 2015). Basic fraction understanding is a noteworthy example because conceptual knowledge of fractions is introduced as early as third grade and their importance for math achievement compounds over time. That is, math skills are often hierarchical by nature—skills from one year are requisite for core instruction in the years to come. When students are missing foundation knowledge (e.g., understanding the magnitude represented by fractions) they are less likely to grasp increasingly complex content (National Mathematics Advisory Panel, 2008; Siegler et al., 2012). This pattern is also evident when examining state test scores for math. In 2022, 40% of Minnesota third grade students failed to pass the state math test. That failure rate increases to 60% by middle school, with over 80% of students from low income households failing to meet state proficiency standards in math (MDE, 2022). Thus, as a Tier 2 intervention program, Math Corps was designed to address skills and standards that fall within and outside grade-level curricula to help build foundational skills that position students to benefit more fully from core instruction.

Relatedly, given the limited time available for tutoring each year, as well as the magnitude of need among students referred for tutoring, there are constraints on the breadth of content that can possibly be supported to mastery via tutoring sessions. In light of these constraints, Math Corps targets student skills in whole number understanding, rational number understanding, and algebraic reasoning. These skill domains are consistently recognized as a critical competency for math development more generally (Gersten et al., 2011; NCTM, 2011), but also specifically for students who lack exposure to high quality math instruction (Flores, 2007). The underlying premise of the Math Corps program is that students who develop a deep, automatic, and flexible understanding of number and operations are in a better position to apply that understanding in other areas of the core curriculum (e.g., measurement and data; geometry).

Math Corps Program Delivery

Math Corps tutors serving students in kindergarten through third grade support students in pairs every day in 20 minute tutoring sessions. The content-focus of each tutoring session follows a logical skill progression that includes grade-level and pre-grade-level whole and rational number concepts. Within each skill group (e.g., addition with regrouping), tutors are able to interchangeably use standardized interventions specific to the unit focus, but variable in regard to the activity. For example, one activity may have a more conceptual focus while another may have a more abstract or applied focus. Varying the nature of support within a similar content focus is by design and centered on developing *Conceptual Understanding, Procedural Fluency,* and *Applied Problem Solving* skills (What Works Clearinghouse [WWC], 2021). Evidence for activities that fall into each of the three categories is largely draw from broad evidence-based principles such as facilitating high opportunities to respond, modeling, and immediate feedback (Carnine et al., 2009; Witzel et al., 2003; Woodward, 2006). However, the formats for specific intervention activities also have direct empirical evidence, such as the Schema-Based activities that compose many of the applied problem solving activities (Jitendra et al., 2006), and number-line based activities across grades (WWC, 2021).

Math Corps tutors serving students in grades four through eight adopt a similar approach; however, the delivery of K-3 support and 4-8 support differs somewhat. More specifically, Math Corps programming for older elementary and middle school students extends its focus to more explicitly address algebraic understanding, tutoring is delivered in groups of two or three, tutoring sessions are between 30 and 45 minutes, and a given intervention may take longer to complete. Yet, the content focus remains largely the same and the support provided within a given skill group includes standardized interventions for conceptual understanding, procedural fluency, and applied problem solving skills.

Finally, all students supported by Math Corps tutors build fact fluency via structured and timed practice distributed across the year. That is, fact fluency is not a “unit” in the Math Corps intervention sequence, but is a general skill that students practice throughout the year.

Data-Based Decision-Making

The use of data to inform key programming decisions is a fundamental component of Math Corps and was a key reference for the Mathway development process. Math Corps tutors use formal and informal data to guide student selection decisions, inform unit or lesson advancement, assess student progress, and determine readiness for exit. At the program-level, researchers use Math

Corps data for grant mandated reporting, impact evaluation, and continuous improvement.

Prior to the Mathway development, tutors supporting students in grades four through eight used STAR Math (Renaissance Learning, 2021) for data-based decision-making with the exception of lesson progress decisions, which were informed by informal brief checks for independent success on the content included in a given lesson or unit. Math Corps expansion to include grades kindergarten through third grade occurred during the same time period as Mathway development and thus Mathway has been used as the primary assessment in grades kindergarten through third grade since the introduction of the program.

In regard to student selection decisions, tutors administer a benchmark assessment to students referred for Math Corps support. Students are eligible for intervention if their benchmark score falls below a benchmark associated with grade-level proficiency. The same criterion-referenced decision is applied seasonally during the winter and spring to guide exit decisions. In regard to unit and/or lesson progression, tutors check for student understanding using an informal assessment of student skill tied directly to a given intervention (4th-8th grade students) or a unit mastery assessment consisting of 6-10 items aligned with unit content (K-3rd grade students). For grant reporting purposes, Math Corps uses student rate of progress as the primary reference point. Prior to the 23-24 academic year, the benchmark for adequate progress is based on exceeding normative growth rates for STAR Math (4th-8th grade) or making seasonal improvement (K-3rd grade).

Important to note, but outside the scope of benchmark assessment, is tutor use of fact fluency assessment data. Across all grades, students complete a 1-minute fact fluency assessment during the same seasonal testing period as the benchmark assessment. The fact fluency assessment is delivered through an online data management system and the benchmark for attainment is equal to 30 problems correct. Tutors use the fact fluency assessment to examine student fact fluency growth as well as to determine when ongoing fact fluency assessment can be discontinued.

Mathway Rationale

As outlined above, Math Corps has historically used STAR Math for screening, response to intervention, and exit decisions, as well as for grant reporting and evaluation. Likewise, STAR Math was the primary outcome of interest in both RCTs evaluating the impact of the program. However, the use of broad-based achievement tests for those decisions limits the program in three important ways. First, broad-based achievement measures only partially align with the content focus of the program. That is, Math Corps is explicitly focused on standards related to whole numbers, rational numbers, and algebraic reasoning whereas broad-based achievement measures assess the full spectrum of grade-level standards. Noting that Math Corps has already demonstrated impacts on broad-based math achievement, aligning the assessment more directly with the content focus of the program will increase the utility of the data for decision-making.

Second, creating an assessment for use within Math Corps creates more autonomy in regard to the content to include, how to present the content, and how to summarize student performance for decision-making. This level of autonomy is critical for ensuring the assessment can adapt to program and community needs, and these kinds of activities are not possible when using published math assessments managed by external parties. Finally, there are significant financial costs associated with external assessments. STAR Math is associated with a $9 cost per student. Given that Math Corps serves thousands of students each year, assessment costs are a significant hurdle for program expansion.

Mathway Content Development Overview

Across all grades, procedures for item writing followed a similar process composed of (1) initial content blueprinting, (2) item writing and review, (3) revision, and (4) field testing. Mathway content development was led by researchers at National Science and Service Collaborative and supported by math researchers, teachers, and Math Corps content leads.

Content Blueprinting

The procedures for content blueprinting included review of national math standards, the Math Corps scope and sequence, and empirical research on foundational math skills. In regard to standards, the research team reviewed national standards associated with whole and rational number understanding and algebraic reasoning, including examples of mastery within each relevant standard. At the same time, the research team reviewed the Math Corps interventions and skill expectations by lesson and unit. Finally, the research team reviewed existing research lines examining foundational math skills that predict future success in math. Collectively, these three sources of information provided the expected skillset across all students within a given-grade (i.e., standard alignment), the expected skillset among students served by the program (i.e., program alignment), and curriculum-independent skills associated with long-term success in math (i.e., empirical review). It is important to note that the development process was not restricted to grade-level content. For example, fourth grade students are expected to multiply fractions (CCSS, 2010), but fraction multiplication assumes foundational understanding of fractions and their relative magnitude, which is introduced in previous grade levels. This is relevant because fourth grade students served by Math Corps are by definition below grade-level and the intent of the program is to position those students to benefit more fully from core instruction. Thus, the scope of the "grade level" Math Corps assessment blueprint is inclusive of foundational content associated with prior grade-levels.

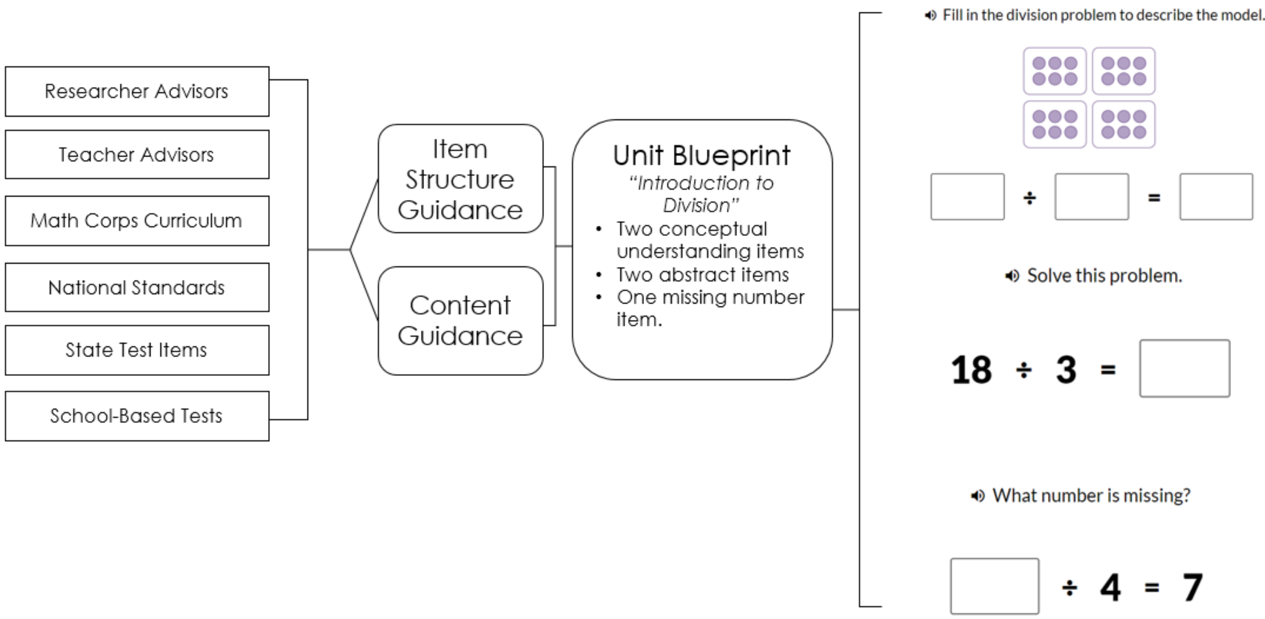
The procedures outlined above helped to identify the content strands to be included on the assessment. In addition, within each content strand, the research team identified the way in which content knowledge would be assessed, including conceptual knowledge, abstract skills, and problem solving skills. For example, for a given content strand (e.g., multi-digit addition) the team considered whether to include conceptual representation items, procedural items, and/or problem-solving items using the Math Corps curriculum, national standards, and empirical research as a guide. The resulting content strands and skill expectations served as a guide for item structure design.

Item Writing and Review

Prior to writing Mathway assessment items, the research team reviewed the structure and nature of existing assessments, including state test items, examples of mastery associated with national standards, and commonly used school-based assessments. This preliminary review was focused primarily on the structure of existing items. That is, in what ways were students asked to demonstrate their skills on existing assessments? In what ways (if any) did those structures relate to the content mastery inherent in the Math Corps program? For example, in cases where students were expected to demonstrate conceptual understanding of whole number subtraction, the research team considered the various ways in which conceptual understanding is discussed in the research literature as well as the ways in which existing research has assessed conceptual understanding of whole number subtraction. Examining the expected performance in the Math Corps interventions was fundamental to aligning the skills targeted by the program and the skills included on the assessment.

In conjunction with the content blueprint, the review of item structures helped to create a blueprint for item writing. For each unit, the research team identified the number and structure of items required. For example, in third grade, Math Corps includes a unit on whole number division. The team engaged in the aforementioned review process and produced an item writing blueprint that included five total items: two conceptual items in which students were required to correctly identify a given division problem among several different semi-abstract representations, two abstract division problems missing the quotient, and one problem in which either the divisor or dividend was missing. In this case, using division to solve word problems was included in a different unit and was thus not included in the blueprint for the *Introduction to Division* Unit. A visual representation of how the general procedures for blueprinting were operationalized for the unit described above is displayed in Figure 1.

Figure 1. Assessment blueprinting procedures operationalized for a third grade unit.



After the creating the assessment blueprint, the research team partnered with a variety of advisors including researchers with expertise in math assessment and intervention (e.g., Anne Foegen [University of Iowa], Sarah Powell [UT-Austin]), Garrett Hall [Florida State University], practicing teachers, and principals to review and provide feedback. The initial review focused primarily on the assessment blueprint and resulted in revisions to better align assessment items with existing research and the needs of school partners. These advisors also engaged in additional review following item writing.

The item writing process for Mathway was primarily led by NSSC staff, Math Corps staff, practicing teachers, and graduate students. The item writing process was iterative and included multiple stages of review. Mathway is delivered three times across the academic year and three items were created to correspond to each season. Across seasons, items differed only in the values included in each item, the structure of items across seasons was identical. During the initial writing stage, contributing writers were assigned related units scaled up through grade-levels. For example, a given writing assignment may have included addition items for kindergarten, scaled upward for similar items in later grades. This approach helped writers ensure minor differences in content expectations across grades were reflected in items and the approach to writing items of a similar nature was consistent. These "draft 1" items were reviewed by item writers assigned to different content strands for clarity, appropriateness, and alignment with the assessment blueprint. The initial item writer then made any revisions necessary to produce "draft 2" items. These items were reviewed by the research team according to the same criteria as the initial review and sent back to the initial item writer. The resulting "draft 3" items were reviewed by external research partners and sent back for a near-final review and revision process completed by the research team.

Final items were documented in a master tracking file and formatted for paper-based delivery during initial field testing. The total number of items included on each test, along with the Math Corps unit associated with those items is displayed in Table 1.

Table 1. Total Number of Mathway Items across Grades and Unit

Grade/Unit	Items
Kindergarten	**23**
Counting & Representing Numbers up to 10	9
Comparing Numbers or Objects	5
Introduction to Addition & Subtraction	5
Word Problem Solving	4
First Grade	**30**
Counting & Representing Numbers	5
Representation of Numbers & Place Value	4
Comparing Numbers or Objects	6
Addition & Subtraction	10
Addition & Subtraction Word Problem Solving	5
Second Grade	**25**
Counting & Representing Numbers	5
Place Value & Comparing Numbers	6
Addition & Subtraction	9
Addition & Subtraction Word Problem Solving	5
Third Grade	**32**
Counting & Representing Numbers	4
Addition & Subtraction	4
Addition & Subtraction Word Problem Solving	3
Introduction to Multiplication	5
Introduction to Division	5
Multiplication & Division Word Problem Solving	4
Introduction to Fractions	3
Comparing Fractions	4
Fourth Grade	**34**
Place Value of Whole Numbers	6
Addition & Subtraction of Whole Numbers	7
Multiplication & Division of Whole Numbers	6
Introduction to Fractions & Decimals	9
Addition & Subtraction of Fractions	6

Grade/Unit	Items
Fifth Grade	**34**
Multiplication & Division of Whole Numbers	6
Introduction to Fractions & Decimals	12
Rational Number: Equivalency	2
Working with Fractions	9
Working with Decimals	5
Sixth Grade	**30**
Whole Number Properties & Adv. Operations	7
Factoring Concepts	3
Fractions, Decimals, & Percents	4
Adding & Subtracting Fractions & Decimals	3
Multiplying & Dividing Fractions & Decimals	4
Ratios & Proportional Relationships	3
Introduction to Algebraic Concepts	6
Seventh Grade	**29**
Whole Number Properties & Adv. Operations	6
Factoring Concepts	5
Positive & Negative Numbers	4
Multiplying & Dividing Fractions & Decimals	4
Introduction to Algebra	5
Ratios & Proportional Relationships	5
Eighth Grade	**31**
Rational Numbers: Properties & Operations	7
Positive & Negative Rational Numbers Operations	4
Adv. Rational & Irrational Number Concepts	11
Proportional Relationship	2
Functions	3
Linear Equations	4

Field Testing and Revision for Online Administration

Mathway items were field tested in the winter and spring of 2021. The purpose of field testing was multifaceted and focused primarily on the relative difficulty of individual items, feasibility, and utility. During the field testing phase, 219 students in grades kindergarten through third grade and 299 students in grades four through eight completed the entire Mathway assessment. Of interest during field testing, was the relative difficulty of each item included on the assessment, the internal consistency of the assessment, and the time taken to complete the assessment. The research team evaluated the item-level performance within each grade, examining the percentage of students who answered each of the items correctly. Descriptive statistics were used to evaluate item difficulty, additional aspects of feasibility (e.g., time taken to complete the assessment), and coefficient alpha was used to obtain estimates of internal consistency. The field testing procedures completed during the 2022 spring benchmark period were used to guide item revisions and item ordering, and the results informed the creation of a discontinue rule for the assessment.

Coefficient alpha estimates across grades during the field test ranged from .75 to .91, providing preliminary evidence for the internal consistency of the Mathway assessment. In regard to item difficulty, all items were originally ordered according to the Math Corps intervention sequence, which is arranged by units. That is, assessment items associated with Unit 1 in any given grade were presented to students before items in Unit 2. In kindergarten through third grade, student performance across the assessment closely mirrored the unit sequence insofar as students tended to accurately respond to items at an increasingly lower rate as the items progressed from the first to last unit. The exception to this pattern in grades K-3 were items written for prior grade-level standards. For example, students in the first grade addition unit are expected to fluently add and subtract within 20 whereas students in kindergarten are expected to develop a strong conceptual understanding of addition and subtraction within 10. Math Corps ensures first grade students have a strong grasp of addition and subtraction by providing support on prior-grade level content and including that content on the associated assessment. However, those items—despite belonging to the addition unit in the grade-level sequence—are markedly easier for students to answer correctly. Thus, the first grade assessment presents those items at the beginning of the assessment, prior to other items associated with the first grade addition unit. In grades K-3, those are the only examples where items are out of sequence relative to the intervention curriculum.

The rate of correct responses across items in grades four through eight were more variable. Conceptually, this is related to the same issue outlined above, albeit at a greater magnitude. Students served by Math Corps in grades four through eight demonstrate much more severe skill deficits, often needing support on content that is two or three grade-levels below the students grade-level. This is reflected in the Math Corps intervention curriculum as well—prior grade-level and current grade-level content is intermixed within units. The variance of difficulty within any given unit for students in grades four through eight was apparent when examining item performance during field testing. In some cases, content from Unit 1 was markedly difficult for students and content from later units was much easier than might be predicted by the Math Corps unit sequence. Thus, the item ordering for Mathway in grades four through eight is less clearly defined by unit progression, with items defined more clearly by individual item difficulty.

Item ordering was a key priority during field testing and revision because of the need to introduce a discontinue rule for administration. The primary purpose of the discontinue rule in this context is related to efficiency—if students struggle on items at the beginning of the test, later items can be reasonably assumed to be incorrect at the benefit of not requiring the student to continue to struggle through the remainder of the test. As an intervention program that supports students who are struggling in math, reducing test anxiety and burnout within a test administration is a critical consideration for Math Corps. There are a variety of methods to estimate the difficulty of items, including those aligned with Item Response Theory (IRT). During field testing, IRT-based analyses and basic descriptive data on item difficulty operationalized as the percentage of students who correctly answered each item produced functionally equivalent conclusions. Thus, the research team adopted the latter approach, using item-level proportion correct data to revise the ordering of the assessment. In addition, the research team evaluated the impact on the association between score on Mathway and scores on school-based universal screening assessments across various discontinue rule scenarios. These scenarios included a full score (all items completed), total score recorded after 3 consecutive incorrect answers, and total score recorded after 4 consecutive incorrect answers. The three scenarios did not impact validity correlations or internal consistency estimates. The research team met with program staff to consider practical issues for testing and based on those sessions, a four-point decision rule was adopted for use.

Mathway Current Use and Technical Evaluation Procedures

Mathway is currently delivered online for students in grades kindergarten through eighth grade. The assessment is delivered three times each year (fall, winter, spring). Tutors assist students in navigating to the Mathway assessment within an online data management system (Math Corps Data Management System; MCDMS) and ensure there are no distractions in the testing space. Students are provided with paper and a writing utensil to solve problems and calculators are not permitted. Students view a small number of practice items before beginning the assessment. Collectively, the practice items expose students to the different item structures in the test (e.g., drag and drop, open response, multiple choice, number line placement). There is no time limit on the test; however, the average time taken to complete the test is approximately 15 minutes. Students begin the test, viewing one item at a time. After advancing to the next item, students are able to access and revise the previous question using a "back" function in the assessment. Students continue answering items until four consecutive incorrect answers are provided. After a test is complete, assessment data are stored in the MCDMS and used to generate item-level, unit-level, and test-level summaries to facilitate data-based decision-making.

Technical Evaluation Implementation Procedures

With the exception of the fall benchmark period for fourth grade through eighth grade students—which included paper-based assessment—the procedures outlined above were in place during the 2022-23 academic year. The order and nature of items did not vary from fall to spring. During the same assessment window in the fall, winter, and spring, all students in grades four through eight completed STAR Math and Mathway. In the context of K-3 Math Corps, data sharing requests were established with schools to obtain universal screening data for students. The research team reviewed the screening data sources and identified aMath (Renaissance Learning, 2022) as the screening assessment used most frequently in schools. Thus, all K-3 student data are included for descriptive analyses and validity data were restricted to partnering schools with aMath data. In addition, a subset of third grade students completed Mathway in spring 2022, concurrently with the Minnesota Comprehensive Assessment (MCA) in Math. An overview of the data sources and implementation schedule related to the current technical evaluation is included in Table 2.

Table 2. Mathway Administration and Criterion Validity Reference across Grades

Grade	Spring 2022	Fall 2022	Winter 2023	Spring 2023	Criterion Reference
Kindergarten		Online	Online	Online	aMath
First Grade		Online	Online	Online	aMath
Second Grade		Online	Online	Online	aMath
Third Grade	Online	Online	Online	Online	aMath; MCA
Fourth Grade		Paper	Online	Online	STAR Math
Fifth Grade		Paper	Online	Online	STAR Math
Sixth Grade		Paper	Online	Online	STAR Math
Seventh Grade		Paper	Online	Online	STAR Math
Eighth Grade		Paper	Online	Online	STAR Math

The sections below detail the technical evaluation of Mathway as implemented during the 2022-23 academic year. Due to differences in the criterion for validity estimates, as well as the current implementation structure of Math Corps, all analyses are generally divided by those pertaining to Mathway in the context of grades K-3 and in the context of grades 4-8.

Mathway in Grades Kindergarten through Third Grade

In this section, we describe our procedures and results for evaluating the psychometric properties of Mathway in grades kindergarten through third grade. For correlation and diagnostic accuracy evaluation, we use Bayesian estimation (McElreath, 2020). This statistical method allows for clearer and more practical interpretation of the uncertainty in estimates, such as the probability of meeting a benchmark based on the student's score. Correlations and diagnostic accuracy can be interpreted the same as traditional methods, with the exception of the uncertainty in the estimates. Rather than using a *p*-value or a confidence interval to determine the statistical significance of a relationship, Bayesian methods provide posterior distributions of estimates, which can be summarized with percentiles and variance measures (e.g., standard deviations). This also means that the distributions of estimates can be interpreted with probabilities. For example, if a correlation is estimated to be .5 with a posterior *SD* of .1, that means there would be a 95% chance the correlation is between .3 and .7. We use this method because it has practical interpretive advantages compared to more traditional methods.

For diagnostic classification analyses, we present area under the curve (AUC) estimates, which are an indicator of how accurately a model detects the binary outcome (e.g., met benchmark or not). We also present post-test probabilities of meeting benchmarks by Mathway Scores (Klingbeil et al., 2021; Hall & Doyle, 2023). Finally, we present odds ratios for diagnostic analyses to depict the change in odds of meeting benchmarks for every .1 increase in Mathway scores.

Student Population

Approximately 3,600 students in grades kindergarten through third grade participated in Math Corps (Kindergarten = 21%, Grade 1 = 25%, Grade 2 = 26%, Grade 3 = 29%) in 59, 79, 78, and 78 sites across all four grades, respectively. Across all four grades, reported race/ethnicity percentages were the following: 36-50% be White, 19-29% Black or African American, 7-12% Hispanic or Latino, 1-2% Asian, 3-4% Multiracial, 2-3% American Indian or Alaska Native, <1% Native Hawaiian or Other Pacific Islander. Fifteen to seventeen percent were reported "Unknown." Across all four grades, 48-56% of students identified as female. Eighty-five percent of students were located in Minnesota, and 15% were from Georgia. These demographics, however, are not necessarily reflected in the psychometric analyses below due to student intervention exit and other sources of missing data.

Table 3. K-3 Demographic Proportions

Variables	Grade			
	KG	1	2	3
Grade Distribution	20%	25%	26%	29%
Gender Distribution				
Female	48%	53%	56%	51%
Male	47%	41%	36%	39%
Unreported Gender	5%	6%	9%	9%
Racial Distribution				
American Indian or Alaska Native	3%	3%	2%	2%
Asian	2%	2%	3%	1%
Black or African American	22%	19%	21%	29%
Hispanic/Latino	9%	7%	10%	12%
Multi-Racial	4%	4%	4%	3%
Native Hawaiian or Other Pacific Islander	0%	0%	-	0%
Unreported Race/Ethnicity	16%	15%	15%	16%
White	44%	50%	42%	36%

Note: Values of 0% are rounded, not actual zeroes.

Assessment Descriptive Data, Reliability, and Item-Level Performance

Descriptive Statistics

Descriptive statistics for K-3 Mathway by season are presented in Table 4 below. Descriptive data are computed separately for each season (i.e., different students may be observed across seasons); however, the general trend is a large increase in the proportion of items correctly answered from one season to the next. Percentiles (in increments of 5%) for scores by grade are presented in the appendix (Table A1). Proportions of correct responses per item (sorted by proportion correct) for Fall, Winter, and Spring of each grade are presented in Figures A1-A3, respectively.

Table 4. K-3 Mathway Descriptive Statistics

Grade	Season	Mean	Median	SD	IQR
KG	Fall	0.32	0.35	0.23	0.43
	Winter	0.60	0.61	0.20	0.26
	Spring	0.81	0.87	0.19	0.22
1	Fall	0.38	0.37	0.18	0.27
	Winter	0.62	0.70	0.25	0.37
	Spring	0.78	0.83	0.18	0.23
2	Fall	0.28	0.28	0.16	0.20
	Winter	0.50	0.48	0.22	0.36
	Spring	0.70	0.76	0.20	0.24
3	Fall	0.14	0.13	0.10	0.09
	Winter	0.36	0.34	0.22	0.34
	Spring	0.58	0.66	0.28	0.47

Note. All values are reported as proportion of items correctly answered. IQR = interquartile range (the range between the 25th and 75th percentiles).

Growth rates by grade are reported in Table 5 below. These predicted proportions and predicted growth rates are estimated from a multilevel Bayesian beta regression model (repeated measures nested in students, students nested in sites), which is designed specifically to model proportions (uninformative Bayesian priors were used). Growth rates are presented as proportions. The values reported in the table below are generated as predictions for hypothetical, future students. For example, using the data in Table 5 below, there is a 50% chance the average kindergarten Fall-Spring growth rate *next year* would fall between .35 and .77 (i.e., .56 +/- .21). Consistent with the descriptive data above, there is a substantial growth in proportion correct across the year, but this varies a trivial amount between seasons (i.e., fall-winter, winter-spring).

Table 5. K-3 Growth Norms

Grade	Measure	Mean	SD	IQR
KG	Fall Proportion	0.23	0.12	0.16
	Winter Proportion	0.52	0.16	0.22
	Spring Proportion	0.79	0.15	0.17
	Fall-Winter Growth	0.29	0.10	0.14
	Winter-Spring Growth	0.26	0.09	0.12
	Fall-Spring Growth	0.56	0.15	0.21
1	Fall Proportion	0.36	0.15	0.22
	Winter Proportion	0.58	0.15	0.21
	Spring Proportion	0.77	0.15	0.19
	Fall-Winter Growth	0.23	0.10	0.14
	Winter-Spring Growth	0.19	0.08	0.10
	Fall-Spring Growth	0.41	0.17	0.23
2	Fall Proportion	0.25	0.12	0.17
	Winter Proportion	0.47	0.16	0.22
	Spring Proportion	0.69	0.17	0.24
	Fall-Winter Growth	0.22	0.10	0.14
	Winter-Spring Growth	0.22	0.09	0.12
	Fall-Spring Growth	0.44	0.18	0.24
3	Fall Proportion	0.14	0.09	0.11
	Winter Proportion	0.34	0.14	0.19
	Spring Proportion	0.59	0.19	0.28
	Fall-Winter Growth	0.19	0.09	0.13
	Winter-Spring Growth	0.26	0.10	0.14
	Fall-Spring Growth	0.45	0.18	0.26

Note: IQR = interquartile range (the range between the 25^{th} and 75^{th} percentiles).

Reliability

Coefficient alpha, a measure of internal consistency (i.e., inter-item correlations), was calculated to assess Mathway K-3 reliability. Table 6 reports the reliability estimates for each grade and season. All reliability estimates fall above typically accepted standards for acceptable reliability (e.g., alpha > .8), although there are several factors to consider in calculating reliability, including the content overlap of items (which would increase the correlations between items and arbitrarily increase reliability) as well as the number of items (adding more items always arbitrarily increases alpha because it increases the between-item variance).

Table 6. K-3 Reliability

Grade and Season	Alpha
Fall K	.89
Fall 1	.86
Fall 2	.84
Fall 3	.84
Winter K	.86
Winter 1	.93
Winter 2	.90
Winter 3	.93
Spring K	.89
Spring 1	.88
Spring 2	.87
Spring 3	.95

Validity and Technical Evidence for Data-Based Decision-Making

We conducted several different analyses to evaluate the predictive and convergent validity of Mathway. These analyses are of particular importance as Mathway is intended to guide student selection and exit decisions, which should have some direct relevance for performance on commonly accepted assessments of proficiency. Our analyses include correlations between Mathway and earlyMath (K-1) or aMath (2-3) as well as diagnostic accuracy analyses between Mathway and early/aMath as well as future Mathway performance. Further, among third grade students, we examine the predictive validity of spring Mathway scores and state test performance.

Correlations were estimated using a multilevel Bayesian regression model with sites as random intercepts (to account for between-site variability). We provide traditional correlation estimates from these models in addition to models that corrected for the restricted range of the sample (i.e., Math Corps students by definition of inclusion have a restricted range of math performance). Correlations are presented in Table 7 below. In general, there is a moderate correlation between Mathway and early/aMath, which provides evidence that Mathway, which focuses on a select subset of math skills, is generally indicative of overall math performance. Further, after correcting for truncation in the data, correlations tended to increase in predictable ways.

Table 7. K-3 Correlations with early/aMath by Grade and Season

Grade	Season	Mean	Mean Corrected	SD	SD Corrected	95% Credible Interval			
						2.5%	2.5% Corrected	97.5%	97.5% Corrected
KG	Fall	0.39	0.68	0.10	0.11	0.20	0.42	0.58	0.85
	Winter	0.43	0.48	0.10	0.11	0.23	0.26	0.64	0.69
	Spring	0.53	0.52	0.09	0.09	0.35	0.34	0.70	0.69
1	Fall	0.35	0.58	0.08	0.10	0.19	0.35	0.52	0.76
	Winter	0.58	0.78	0.09	0.07	0.41	0.63	0.74	0.90
	Spring	0.38	0.56	0.09	0.10	0.21	0.34	0.55	0.74
2	Fall	0.30	0.48	0.07	0.09	0.17	0.29	0.43	0.64
	Winter	0.29	0.49	0.06	0.08	0.17	0.32	0.40	0.64
	Spring	0.42	0.65	0.06	0.07	0.30	0.51	0.54	0.77
3	Fall	0.34	0.40	0.10	0.11	0.15	0.19	0.53	0.61
	Winter	0.37	0.32	0.07	0.07	0.22	0.19	0.51	0.45
	Spring	0.33	0.28	0.07	0.06	0.19	0.16	0.47	0.41

Note: Credible intervals can be interpreted as actual probabilities (i.e., there is a 95% chance the estimate falls between the values of the 2.5 and 97.5 percentiles).

In Table 8 below, we also present correlations of Mathway across seasons by grade. These correlations are small to moderate in size. Most correlations (except Winter-Spring K) are substantially different than 0 (even if the 95% interval covers 0 to a small degree).

Table 8. K-3 Mathway Correlations Across Seasons

Grade	Season	Mean	SD	95% Credible Interval	
				2.5%	97.5%
KG	Fall-Winter	0.35	0.11	0.14	0.56
	Winter-Spring	0.06	0.13	-0.19	0.32
	Fall-Spring	0.21	0.12	-0.03	0.45
1	Fall-Winter	0.24	0.11	0.02	0.45
	Winter-Spring	0.31	0.09	0.12	0.49
	Fall-Spring	0.11	0.10	-0.08	0.32
2	Fall-Winter	0.32	0.04	0.24	0.39
	Winter-Spring	0.47	0.04	0.40	0.55
	Fall-Spring	0.31	0.05	0.21	0.40
3	Fall-Winter	0.35	0.04	0.27	0.43
	Winter-Spring	0.43	0.04	0.36	0.51
	Fall-Spring	0.19	0.04	0.11	0.28

Note. Credible intervals can be interpreted as actual probabilities (i.e., there is a 95% chance the estimate falls between the values of the 2.5 and 97.5 percentiles).

Accuracy for Early/aMath

To further assess convergent validity, we assessed the accuracy of Mathway for predicting whether students met benchmarks on early/aMath (as set by Math Corps). We used Bayesian multilevel logistic regression (with random intercepts for sites) to model the predictive accuracy of Mathway in each grade.

We find that area under the curve values, which is one way to assess the predictive accuracy of a measure, ranged from .62 to .80 across grades and seasons (a value of .5 indicates chance). Note that the observed AUC values are artificially small in this case due to the truncated nature of the data. Unlike correlations, there was not a technically defensible way to adjust the AUC values for a truncated sample. Across grade levels and seasons, we found that a performance increase of 10% (i.e., increasing a score by +.1) on Mathway was associated with being 1.5-2x as likely to meet the seasonal benchmark. These effects were reliably different from zero across grades and seasons.

In Table A2, we provide probabilities of meeting early/aMath benchmarks within each season by Mathway scores (in increments of .1). These estimates are conceptually similar to post-test probabilities obtained through interval likelihood ratios (Klingbeil et al., 2021) and are a fundamental reference point for making Math Corps exit decisions as they represent the probability of student success on broad-based measures of math achievement. To help inform the practical use of the assessment, we examined the Mathway scores and associated probabilities to identify a meaningful Mathway benchmark. More specifically, we used the spring Mathway scores and associated probabilities of reaching proficiency on early/aMath to identify a spring cut score that could be used uniformly across seasons as a Mathway exit criterion.

As discussed previously, the probability of reaching grade-level expectations on earlyMath/aMath increased as Mathway scores increased. In general, a Mathway score between 70% and 80% correct was associated with greater than 50% probability of meeting grade-level benchmarks. Given the intent of the Math Corps program to position students for success in their classroom, the use of a 50% probability benchmark set at approximately a 75% correct score on Mathway is likely an appropriate benchmark as it indicates (1) near mastery of Math Corps content, (2) readiness for grade-level classroom instruction, and (3) moderate probability of meeting grade-level expectations. It is important to note that the interpretation of exit decisions in this context is not solely focused on ensuring students will be proficient on grade-level tests, it is also intended to signify when students can fully benefit from core instruction without the need for additional support on foundational skills. The use of a 75% correct benchmark on Mathway ensures confidence that a given student has mastered a majority of foundational content, is prepared to more fully benefit from core instruction, and has greater than 50% chance of scoring at grade-level on a comprehensive math assessment at the time of exit. The impact of tutoring is thus focused on increasing core math skills and increasing student probability of success. For example, during the 2022-2023 academic year, the average third grade score on Mathway increased from 14% in the fall to 59% in the spring, which is equivalent to approximately a 24% increase in the probability of success on earlyMath/aMath (using Table A2 in the Appendix as a reference).

Accuracy for 2022 Grade 3 State Test Data

To examine the association between Mathway and state test outcomes, we used Grade 3 data from Spring of 2022. We find that the correlation between third grade students Mathway and state test scores in Spring 2022 was .59 (95% credible interval: 0.42-0.72)—nearly double the size of the correlation with aMath in third grade. We also find that an increase of 10% on Spring 2022 Mathway scores predicted that students were 1.4 times less likely to be at risk. The risk probabilities for third grade students scoring approximately 75% on the Mathway assessment were generally commensurate with those observed on earlyMath/aMath, which lends additional evidence for the use of a 75% uniform benchmark for student performance on Mathway.

Between-Season Accuracy

We also examined the accuracy of Fall and Winter Mathway in predicting whether students in Spring scored 75% or higher on Mathway. We established using 75% as the "target" score for Mathway because this was a score that provided a reliable (but attainable) indication that a student was better positioned to meet the earlyMath/aMath spring benchmark (see Table A2). Table A3 provides post-test probabilities of scoring >75% by Mathway score.

Across grades, we find that the area under the curve of Fall Mathway for predicting Spring performance >75% ranged from .59 to .66 (low accuracy). This is generally to be expected given that student performance is expected to significantly change over the course of the year, such that Fall performance is not a strong indicator of where a student performs in the spring given their response to intervention. In addition, average fall performance was markedly and uniformly low among K-3 students, which restricts the potential for explanatory power. An increase in Mathway score of .1 in Fall performance related to being between 1.2-1.7x times more likely to score >75% in the Spring (these estimates are reliably different from 0).

Winter Mathway was more accurate for predicting Spring performance >75%, with area under the curve values between .70 and .76. A 10% increase in Winter performance related to being between 1.3-1.7x times more likely to score >75% in the Spring (these estimates are reliably different from 0). Altogether, these results suggest that as students increase their Mathway scores as a function of receiving intervention, they increase their likelihood of meeting Spring benchmarks on criterion measures.

Mathway in Grades Four through Eight

In this section, we describe our procedures and results for evaluating the psychometric properties of Mathway. For correlation and diagnostic accuracy evaluation, we use Bayesian estimation (McElreath, 2020). This statistical method allows a clearer and more practical interpretation of the uncertainty in estimates, such as the probability of meeting a benchmark based on the student's score. Correlations and diagnostic accuracy can be interpreted the same as traditional methods, with the exception of the uncertainty in the estimates. Rather than using a *p*-value or a confidence interval to determine the statistical significance of a relationship, Bayesian methods provide posterior distributions of estimates, which can be summarized with percentiles and variance measures (e.g., standard deviations). This also means that the distributions of estimates can be interpreted with probabilities. For example, if a correlation is estimated to be .5 with a posterior *SD* of .1, that means there would be a 95% chance the correlation is between .3 and .7. We use this method because it has practical interpretive advantages compared to more traditional methods.

For diagnostic classification analyses, we present area under the curve (AUC) estimates, which are an indicator of how accurately a model detects the binary outcome (e.g., met benchmark or not). We also present post-test probabilities of meeting benchmarks by Mathway Scores (Hall & Doyle, 2023; Klingbeil et al., 2021).

Student Population

Approximately 2,700 students in grades 4-8 completed Mathway across the fall, winter, or spring (Grade 4 = 35%, Grade 5 = 38%, Grade 6 = 13%, Grade 7 = 8%, Grade 8 = 6%) in 90, 95, 37, 25, and 23 sites across all five grades, respectively. Across grades, reported race/ethnicity percentages were the following: 22-40% be White, 10-22% Black or African American, 9-16% Hispanic or Latino, 2-5% Asian, 3-5% Multiracial, 1-2% American Indian or Alaska Native, <1% Native Hawaiian or Other Pacific Islander. Fourteen to twenty percent were reported

"Unknown." Across all four grades, 43-49% of students were reported to be female, with 7-19% "unknown." Ninety-five percent of students were located in Minnesota, with the rest located in California, Colorado, Georgia, Michigan, and Mississippi. These demographics, however, are not necessarily reflected in the psychometric analyses below due to student intervention exit and other sources of missing data.

Table 9. Grades 4-8 Demographic Proportions

Variables	Grade				
	4	5	6	7	8
Grade Distribution	0.35	0.37	0.13	0.08	0.06
Gender Distribution					
Female	0.48	0.44	0.44	0.49	0.43
Male	0.32	0.37	0.39	0.44	0.46
Unreported Gender	0.20	0.19	0.17	0.08	0.11
Racial Distribution					
American Indian or Alaska Native	0.02	0.01	0.01	0.00	0.01
Asian	0.03	0.05	0.03	0.02	0.04
Black or African American	0.10	0.12	0.11	0.18	0.22
Hispanic/Latino	0.09	0.09	0.11	0.14	0.16
Multi-Racial	0.04	0.05	0.04	0.03	None
Native Hawaiian or Other Pacific Islander	None	0.00	None	0.00	0.01
Unreported Race/Ethnicity	0.32	0.34	0.35	0.24	0.35
White	0.40	0.33	0.35	0.38	0.22

Note. Values of 0.00 are rounded, not actual zeroes.

Descriptive Data, Reliability, and Item-Level Performance

Descriptive Statistics

Descriptive statistics for 4-8 Mathway by season are presented in Table 10 below. Descriptives are computed separately for each season (i.e., different students may be observed across seasons), but the general trend is a consistent increase in the proportion of items correctly answered. Percentiles (in increments of 5%) for scores by grade are presented in Table A4. Proportions of correct responses by item (sorted by proportion correct) for Fall, Winter, and Spring of each grade are presented in Figures A4-A6, respectively.

Table 10. 4-8 Mathway Descriptives

Grade	Season	Mean	Median	SD	IQR
4	Fall	0.19	0.18	0.09	0.09
	Winter	0.29	0.29	0.15	0.21
	Spring	0.45	0.47	0.19	0.29
5	Fall	0.11	0.09	0.10	0.12
	Winter	0.19	0.15	0.17	0.29
	Spring	0.25	0.24	0.15	0.24
6	Fall	0.13	0.10	0.10	0.13
	Winter	0.25	0.23	0.18	0.23
	Spring	0.31	0.30	0.22	0.27
7	Fall	0.15	0.14	0.15	0.17
	Winter	0.23	0.21	0.16	0.27
	Spring	0.29	0.24	0.20	0.30
8	Fall	0.11	0.07	0.12	0.12
	Winter	0.17	0.12	0.14	0.21
	Spring	0.19	0.12	0.16	0.19

Note. IQR=interquartile range

Growth rates by grade are reported below. These predicted proportions and predicted growth rates are estimated from a multilevel Bayesian beta regression model (repeated measures nested in students, students nested in sites), which is designed specifically to model proportions (uninformative Bayesian priors were used). Growth rates are presented as proportions. The values reported in the table below are generated as predictions for hypothetical, future students (i.e., these are not the predictions for the students observed in the current data; rather, these are predictions for what future data may look like based on the available data in the current Mathway K-3 data). For example, from the data below, we could say that there is a 50% chance the average Grade 4 Fall-Spring growth rate *next year* would fall between .11 and .43 (i.e., .27 +/- .16). Consistent with the descriptive data above, there is growth in proportion correct across the year, and this varies a trivial amount between seasons (i.e., fall-winter, winter-spring).

Table 11. Mathway Growth Norms for Students in Grades 4-8

Grade	Measure	Mean	SD	IQR
4	Fall Proportion	0.18	0.11	0.13
	Winter Proportion	0.29	0.13	0.18
	Spring Proportion	0.45	0.15	0.22
	Fall-Winter Growth	0.12	0.05	0.07
	Winter-Spring Growth	0.16	0.07	0.09
	Fall-Spring Growth	0.27	0.11	0.16
5	Fall Proportion	0.11	0.08	0.09
	Winter Proportion	0.17	0.09	0.11
	Spring Proportion	0.25	0.12	0.16
	Fall-Winter Growth	0.06	0.04	0.04
	Winter-Spring Growth	0.08	0.05	0.07
	Fall-Spring Growth	0.14	0.09	0.12
6	Fall Proportion	0.13	0.09	0.10
	Winter Proportion	0.20	0.10	0.13
	Spring Proportion	0.29	0.13	0.18
	Fall-Winter Growth	0.06	0.04	0.05
	Winter-Spring Growth	0.10	0.06	0.08
	Fall-Spring Growth	0.16	0.10	0.13
7	Fall Proportion	0.12	0.08	0.09
	Winter Proportion	0.17	0.09	0.12
	Spring Proportion	0.26	0.12	0.17
	Fall-Winter Growth	0.06	0.04	0.05
	Winter-Spring Growth	0.09	0.06	0.08
	Fall-Spring Growth	0.15	0.09	0.13
8	Fall Proportion	0.09	0.06	0.07
	Winter Proportion	0.12	0.07	0.08
	Spring Proportion	0.18	0.09	0.12
	Fall-Winter Growth	0.04	0.03	0.03
	Winter-Spring Growth	0.06	0.05	0.06
	Fall-Spring Growth	0.09	0.07	0.09

Note. IQR = interquartile range, which is the range between the 25th and 75th percentiles.

Reliability

Coefficient alpha, a measure of internal consistency (i.e., inter-item correlations), was calculated to assess Mathway 4-8 reliability. Table 13 reports the reliability estimates for each grade and season. All reliability estimates fall above typically accepted standards for reliability (e.g., alpha > .8), although there are several factors to consider in calculating reliability, including the content overlap of items (which would increase the correlations between items and arbitrarily increase reliability) as well as the number of items (adding more items always arbitrarily increases alpha because it increases the between-item variance).

Table 12. 4-8 Mathway Reliability

Grade and Season	Alpha
Fall 4	0.77
Fall 5	0.84
Fall 6	0.77
Fall 7	0.89
Fall 8	0.91
Winter 4	0.87
Winter 5	0.91
Winter 6	0.88
Winter 7	0.87
Winter 8	0.88
Spring 4	0.90
Spring 5	0.87
Spring 6	0.91
Spring 7	0.90
Spring 8	0.89

Validity and Technical Evidence for Data-Based Decision-Making

We conducted several different analyses to evaluate the predictive and convergent validity of Mathway. Our analyses include correlations between Mathway and Star Math as well as diagnostic accuracy analyses between Mathway and Star Math as well as future Mathway performance.

Correlations were estimated using a multilevel Bayesian regression model with sites as random intercepts (to account for between-site variability). We provide traditional correlation estimates from these models in addition to models that corrected for the restricted range of the sample (i.e., Math Corps students by definition of inclusion have a restricted range of math performance). Correlations are presented in Table 14 below. We generally find that there is a moderate correlation between Mathway and Star Math, which provides evidence that Mathway—which measures only subset of grade-level math skills—offers meaningful predictive value for student performance on comprehensive measures of math achievement.

Table 13. Grade 4-8 Mathway Correlations with Star Math

Grade	Season	Mean	Mean Corrected	SD	SD Corrected	95% Credible Interval			
						2.5%	2.5% Corrected	97.5%	97.5% Corrected
4	Fall	0.36	0.60	0.05	0.06	0.27	0.48	0.29	0.51
	Winter	0.37	0.52	0.03	0.04	0.30	0.44	0.32	0.46
	Spring	0.53	0.69	0.03	0.03	0.47	0.63	0.48	0.65
5	Fall	0.41	0.65	0.05	0.05	0.32	0.53	0.34	0.57
	Winter	0.45	0.68	0.03	0.04	0.38	0.60	0.39	0.62
	Spring	0.59	0.76	0.03	0.03	0.53	0.70	0.55	0.72
6	Fall	0.35	0.57	0.09	0.12	0.17	0.31	0.22	0.39
	Winter	0.48	0.63	0.06	0.06	0.36	0.50	0.39	0.54
	Spring	0.57	0.69	0.06	0.05	0.45	0.58	0.48	0.61
7	Fall	0.24	0.58	0.07	0.13	0.10	0.29	0.14	0.38
	Winter	0.51	0.72	0.07	0.06	0.37	0.59	0.41	0.63
	Spring	0.32	0.43	0.08	0.10	0.16	0.23	0.20	0.28
8	Fall	0.17	0.35	0.08	0.16	0.00	0.01	0.05	0.10
	Winter	0.51	0.71	0.08	0.08	0.34	0.53	0.39	0.59
	Spring	0.45	0.64	0.09	0.10	0.26	0.42	0.31	0.49

Note: Credible intervals can be interpreted as actual probabilities (i.e., there is a 95% chance the estimate falls between the values of the 2.5 and 97.5 percentiles).

In Table 14 below, we also present correlations of Mathway across seasons by grade. These correlations are small to moderate in size, though some are not substantially different from 0 (smaller correlations with greater estimated uncertainty). As was the case with the K-3 Mathway assessment, fall scores were uniformly and markedly low, which restricts the magnitude of observed correlations due to a restricted range.

Table 14. 4-8 Mathway Correlations Across Seasons

Grade	Season	Mean	SD	95% Credible Interval 2.5%	97.5%
4	Fall-Winter	0.25	0.07	0.10	0.40
	Winter-Spring	0.50	0.04	0.42	0.58
	Fall-Spring	0.22	0.08	0.06	0.38
5	Fall-Winter	0.47	0.07	0.33	0.61
	Winter-Spring	0.53	0.04	0.45	0.61
	Fall-Spring	0.61	0.09	0.43	0.79
6	Fall-Winter	0.16	0.13	-0.11	0.41
	Winter-Spring	0.46	0.09	0.28	0.65
	Fall-Spring	-0.11	0.17	-0.43	0.22
7	Fall-Winter	0.26	0.14	0.01	0.53
	Winter-Spring	0.41	0.09	0.23	0.59
	Fall-Spring	0.30	0.16	-0.03	0.62
8	Fall-Winter	0.47	0.16	0.14	0.79
	Winter-Spring	0.46	0.12	0.24	0.69
	Fall-Spring	0.15	0.22	-0.30	0.58

Note: Credible intervals can be interpreted as actual probabilities (i.e., there is a 95% chance the estimate falls between the values of the 2.5 and 97.5 percentiles).

Accuracy for Star Math

To further assess convergent validity, we assessed the accuracy of Mathway for predicting whether students met benchmarks on Star Math. We used Bayesian multilevel logistic regression (with random intercepts for sites) to model the predictive accuracy of Mathway in each grade.

We find that area under the curve values, which is one way to assess the predictive accuracy of a measure, ranged from .63 to .86 across grades and seasons (a value of .5 indicates chance). Across grade levels and seasons, we found that a performance increase of 10% on Mathway was associated with being 1.6-3x more likely to meet each seasonal benchmark (however, these estimates in Grades 6-7 for the fall are not reliably different than 0, and no students in Grade 8 met the fall benchmark). Altogether, these values indicate moderate accuracy of Mathway for predicting meeting the Star proficiency benchmark in each season.

In Table A5, we provide probabilities of meeting Star Math seasonal benchmarks based on Mathway scores (in increments of .1). These estimates are conceptually similar to post-test probabilities obtained through interval likelihood ratios (Klingbeil et al.,

2021) and are a fundamental reference point for making Math Corps exit decisions as they represent the probability of student success on broad-based measures of math achievement. To help inform the practical use of the assessment, we examined the Mathway scores and associated probabilities to identify a meaningful Mathway benchmark. More specifically, we used the spring Mathway scores and associated probabilities of reaching proficiency on Star Math to identify a spring cut score that could be used uniformly across seasons as a Mathway exit criterion. Barring no major discrepancies in predictive value across grades, a uniform benchmark across season affords a great deal of simplicity for exit decision-making for tutors who serve students across different grades.

As discussed previously, the probability of reaching grade-level expectations on Star Math increased as Mathway scores increased. In general, a Mathway score between 70% and 80% correct was associated with greater than 50% probability of meeting grade-level benchmarks. Given the intent of the Math Corps program to position students for success in their classroom, the use of a 50% probability benchmark set at approximately a 75% correct score on Mathway is likely an appropriate benchmark as it indicates (1) near mastery of Math Corps content, (2) readiness for grade-level classroom instruction, and (3) moderate probability of meeting grade-level expectations. It is important to note that the interpretation of exit decisions in this context is not solely focused on ensuring students will be proficient on grade-level tests, it is also intended to signify when students can fully benefit from core instruction without the need for additional support on foundational skills.

The use of a 75% correct benchmark on Mathway ensures confidence that a given student has mastered a majority of foundational content, is prepared to more fully benefit from core instruction, and has greater than 50% chance of scoring at grade-level on a comprehensive math assessment at the time of exit. The impact of tutoring is thus focused on increasing core math skills and increasing student probability of success. For example, during the 2022-2023 academic year, the average fourth grade score on Mathway increased from 18% in the fall to 45% correct in the spring, which is equivalent to approximately a 22% increase in the probability of meeting the proficiency criterion on Star Math (see Table A5).

Between-Season Accuracy

We also examined the accuracy of Fall and Winter Mathway scores in predicting whether students scored 75% or higher on Mathway in the spring (the aforementioned benchmark score). We established using 75% as the "target" score for Mathway because this was a score that provided a reliable indication that a student was likely to meet the Star spring benchmark (see Table A5).

Across grades, we find that the area under the curve of Fall Mathway for predicting Spring performance >75% ranged from .61 to .78 (low to moderate accuracy). This is generally to be expected given that student performance is expected to significantly change over the course of the year, such that Fall performance is not a strong indicator of where a student performs in the spring given their response to intervention and overall uniform low performance in the fall. A 10% increase in Fall performance related to being between 1.9-2.1x times more likely to score >75% in the Spring, but these estimates are highly uncertain and should be interpreted with caution (these estimates do not include Grade 6, which was too unreliable to interpret, or Grade 5, in which no students with Mathway scores met the the 75% criterion in Spring).

Winter Mathway scores were more accurate for predicting Spring performance >75%, with AUC values between .64 and .96 (low to moderate accuracy). A 10% increase in Winter performance related to being between 1.4-3.2x times more likely to score >75% in the Spring, though these results from Grades 6-7 are not reliably different from 0. Altogether, these results suggest that as students increase their Mathway scores as a function of receiving intervention, they increase their likelihood of meeting Spring benchmarks on criterion measures.

In general terms, the between-season accuracy of Mathway (in terms of predicting the Spring 75% criterion) is quite low in certain cases, and this is primarily due to the small number of students that have Fall or Winter Mathway data (which in some grades has a very truncated distribution in the Fall) and also score 75% or above in the Spring. Although reaching a score of 75% in the Spring is a reliable indicator of meeting the STAR Benchmark, it may be difficult to predict across seasons who will actually reach a score of 75%. This may happen in part due to students' response to intervention over time (i.e., where a student scores in the fall is not necessarily an indicator of where they score in the spring if they respond to the intervention). The base rate of students scoring >75% in the Spring is also very low in some cases, and it is also likely to remain low given the distribution of scores in later grades in which students do not score as high on average as they do in Grades 4 or 5 (or, likewise, in Grades K-3).

Implications for Practice

Data-based decision-making is a core feature of the Math Corps program. In the context of tri-annual screening, Math Corps uses data to identify students for support, monitor response to intervention, and make exit decisions. Data in the current report provide guidance for the ways in which the

Mathway assessment can be used by the program to make those decisions.

During the 2023-2024 program year, the Math Corps program team and Research team at ServeMinnesota will finalize the recommendations for practice, which will include the associated impact on data displays for interpretation in the Math Corps data management system. These recommendations will be incorporated into training and coaching materials.

For the purposes of decision-making during the 2023-2024 program year, the program and research teams used the current report to guide recommendations for student selection and exit decisions.

In regard to identifying students for support, it is likely that nearly all students who are selected for Mathway screening are good candidates for Math Corps. This is largely due to the fact that students selected to participate in Math Corps screening have already been identified based on existing data evidencing a need for intervention. Indeed, the Mathway data indicate that a majority of students score less than 20% correct in the fall, with no student scores above 50% correct. Given that all students identified for Math Corps support in the fall are likely to benefit from the program, program staff and internal coaches are encouraged to use Mathway data and other important school-based information to make a local decision regarding the students to enroll in support. Students with lower scores are likely to have larger skill gaps and might be expected to remain in tutoring for the duration of the year whereas students who begin the year with larger Mathway scores are more likely to exit intervention. In all cases, students who score below 75% correct on Mathway can be considered an appropriate fit for Math Corps support.

In regard to exit decisions, the probabilities of meeting grade-level expectations associated with 10% increments on Mathway can be used to guide the adoption and use of a single exit criterion across seasons. In the current report, we outline the use of a 75% correct criterion as it affords confidence that students have mastered a majority of foundational skills included in the Math Corps curriculum, are positioned to benefit more directly from core instruction, and have a probability of meeting grade-level benchmarks that is greater than or equal to approximately 50%. However, there remains considerable uncertainty in how Mathway earlier in the year predicts meeting the 75% criterion in Spring, particularly in later grades. Although a score of >75% reliably predicts meeting the STAR benchmark, there are in fact few students that score >75% in Spring. This means it is at times difficult to predict who will or will not meet that criteria. More data is needed to obtain more reliable predictions.

References

Carnine, D. W., Silbert, J., Kame'ennui, E. J., & Tarver, S. G. (2009). Direct instruction reading. 5th.

Codding, R. S., Nelson, P. M., Parker, D. C., Edmunds, R., & Klaft, J. (2022). Examining the impact of a tutoring program implemented with community support on math proficiency and growth. *Journal of School Psychology*, *90*, 82-93.

Flores, A. (2007). Examining disparities in mathematics education: Achievement gap or opportunity gap?. *The High School Journal*, *91*(1), 29-42.

Fuchs, L. S., Bucka, N., Clarke, B., Dougherty, B., Jordan, N. C., Karp, K. S., ... & Morgan, S. (2021). Assisting Students Struggling with Mathematics: Intervention in the Elementary Grades. Educator's Practice Guide. WWC 2021006. *What Works Clearinghouse*.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). Assisting students struggling with mathematics: Response to intervention (RtI) for elementary and middle schools. *IES National Center for Education Evaluation Practice Guide*.

Hall, G., & Doyle, E. (2023). Back to Bayes-ics: Improving Universal Screening Decisions by Quantifying Uncertainty.

Jitendra, A. K., Nelson, G., Pulles, S. M., Kiss, A. J., & Houseworth, J. (2016). Is mathematical representation of problems an evidence-based strategy for students with mathematics difficulties?. *Exceptional Children*, *83*(1), 8-25.

Klingbeil, D. A., Van Norman, E. R., & Nelson, P. M. (2021). Using interval likelihood ratios in gated screening: A direct replication study. *Assessment for Effective Intervention*, *47*(1), 14-25.

Lortie-Forgues, H., Tian, J., & Siegler, R. S. (2015). Why is learning fraction and decimal arithmetic so difficult?. *Developmental Review*, *38*, 201-221.

National Mathematics Advisory Panel. (2008). The final report of the national mathematics advisory panel. Washington, DC: Department of Education.

Nelson, P. M., Parker, D. C., & Van Norman, E. R. (2018). Subskill mastery among elementary and middle school students at risk in mathematics. *Psychology in the Schools*, *55*(6), 722-736.

Nelson, P. M., Pulles, S. M., Parker, D. C., & Klaft, J. (2020). Implementation fidelity for math intervention: Basic quality ratings to supplement adherence. *School Psychology*, *35*(1), 72.

Nelson, P. M., Van Norman, E. R., Parker, D. C., & Cormier, D. C. (2019). An examination of interventionist implementation fidelity and content knowledge as predictors of math intervention effectiveness. *Journal of Applied School Psychology*, *35*(3), 234-256.

Parker, D. C., Nelson, P. M., Zaslofsky, A. F., Kanive, R., Foegen, A., Kaiser, P., & Heisted, D. (2019). Evaluation of a math intervention program implemented with community support. *Journal of Research on Educational Effectiveness*, *12*(3), 391-412.

Renaissance Learning, Inc. (2021). Star Assessments for Math technical manual. https://vdocuments.site/star-assessments-for-math-technical-manual-rp-introductiondesign-of-star.html?page=1

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., ... & Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological science*, *23*(7), 691-697.

Spaull, N., & Kotze, J. (2015). Starting behind and staying behind in South Africa: The case of insurmountable learning deficits in mathematics. *International Journal of Educational Development*, *41*, 13-24.

Van Norman, E. R., Nelson, P. M., & Parker, D. C. (2017). Technical adequacy of growth estimates from a computer adaptive test: Implications for progress monitoring. *School Psychology Quarterly*, *32*(3), 379.

Witzel, B. S., Mercer, C. D., & Miller, M. D. (2003). Teaching algebra to students with learning difficulties: An investigation of an explicit instruction model. *Learning Disabilities Research & Practice*, *18*(2), 121-131.

Woodward, J. (2006). Developing automaticity in multiplication facts: Integrating strategy instruction with timed practice drills. *Learning Disability Quarterly*, *29*(4), 269-289.

Appendices

Table A1. K-3 Score Percentiles by Grade

Season	Percentile	Mathway Score			
		Grade K	Grade 1	Grade 2	Grade 3
Fall	0.00	0.00	0.00	0.00	0.00
	0.05	0.00	0.12	0.04	0.03
	0.10	0.00	0.17	0.08	0.03
	0.15	0.04	0.20	0.12	0.06
	0.20	0.04	0.23	0.16	0.06
	0.25	0.04	0.23	0.16	0.09
	0.30	0.13	0.23	0.20	0.09
	0.35	0.22	0.27	0.20	0.09
	0.40	0.26	0.30	0.24	0.13
	0.45	0.30	0.33	0.24	0.13
	0.50	0.35	0.37	0.28	0.13
	0.55	0.39	0.40	0.28	0.16
	0.60	0.39	0.43	0.28	0.16
	0.65	0.43	0.47	0.32	0.16
	0.70	0.48	0.47	0.32	0.16
	0.75	0.48	0.50	0.36	0.19
	0.80	0.52	0.53	0.40	0.19
	0.85	0.57	0.57	0.44	0.19
	0.90	0.61	0.63	0.48	0.22
	0.95	0.70	0.68	0.60	0.31
	1.00	0.87	0.87	0.96	0.91
Winter	0.00	0.00	0.03	0.00	0.00
	0.05	0.18	0.17	0.19	0.06
	0.10	0.39	0.21	0.24	0.09
	0.15	0.43	0.27	0.28	0.13
	0.20	0.48	0.37	0.32	0.16
	0.25	0.48	0.43	0.32	0.19
	0.30	0.52	0.50	0.36	0.19
	0.35	0.57	0.57	0.40	0.22
	0.40	0.57	0.60	0.44	0.25
	0.45	0.57	0.67	0.48	0.31
	0.50	0.61	0.70	0.48	0.34
	0.55	0.64	0.70	0.52	0.38
	0.60	0.65	0.73	0.56	0.41
	0.65	0.70	0.77	0.60	0.44

	0.70	0.70	0.80	0.64	0.47
	0.75	0.74	0.80	0.68	0.53
	0.80	0.78	0.83	0.72	0.56
	0.85	0.83	0.87	0.76	0.63
	0.90	0.83	0.90	0.80	0.69
	0.95	0.91	0.93	0.88	0.78
	1.00	1.00	1.00	1.00	0.97
Spring	0.00	0.00	0.00	0.00	0.00
	0.05	0.49	0.43	0.32	0.13
	0.10	0.57	0.50	0.40	0.16
	0.15	0.65	0.57	0.48	0.22
	0.20	0.74	0.63	0.56	0.25
	0.25	0.74	0.67	0.60	0.34
	0.30	0.78	0.70	0.64	0.41
	0.35	0.80	0.77	0.67	0.47
	0.40	0.83	0.77	0.68	0.53
	0.45	0.87	0.80	0.72	0.59
	0.50	0.87	0.83	0.76	0.66
	0.55	0.87	0.83	0.76	0.69
	0.60	0.91	0.87	0.80	0.72
	0.65	0.91	0.87	0.80	0.75
	0.70	0.91	0.90	0.84	0.78
	0.75	0.96	0.90	0.84	0.81
	0.80	0.96	0.93	0.87	0.84
	0.85	0.96	0.93	0.88	0.88
	0.90	1.00	0.97	0.92	0.91
	0.95	1.00	1.00	0.96	0.97
	1.00	1.00	1.00	1.00	1.00

Table A2. Predicted early/aMath Benchmark Probabilities by Mathway Score

Mathway Score	Grade	Odds Ratio (95% Interval)	Mean Probability	95% Credible Interval 2.50%	95% Credible Interval 97.5
Fall					
0.0	**KG**	1.34 (0.97-1.92)	0.04	0.00	0.14
0.1			0.05	0.00	0.15
0.2			0.06	0.01	0.17
0.3			0.08	0.01	0.20
0.4			0.11	0.01	0.25
0.5			0.14	0.02	0.33
0.6			0.18	0.02	0.43
0.7			0.22	0.03	0.55
0.8			0.28	0.03	0.67
0.9			0.33	0.03	0.78
1.0			0.39	0.03	0.87
0.0	**1**	1.31 (1.00-1.77)	0.06	0.01	0.16
0.1			0.08	0.02	0.17
0.2			0.09	0.03	0.18
0.3			0.12	0.05	0.21
0.4			0.15	0.06	0.24
0.5			0.19	0.08	0.31
0.6			0.23	0.09	0.40
0.7			0.29	0.09	0.52
0.8			0.34	0.10	0.64
0.9			0.41	0.10	0.75
1.0			0.47	0.10	0.84
0.0	**2**	1.59 (1.17-2.16)	0.04	0.01	0.12
0.1			0.06	0.01	0.15
0.2			0.09	0.03	0.19
0.3			0.13	0.04	0.26
0.4			0.19	0.07	0.36
0.5			0.27	0.10	0.49
0.6			0.37	0.14	0.64
0.7			0.47	0.17	0.78
0.8			0.57	0.21	0.88
0.9			0.66	0.25	0.94
1.0			0.73	0.29	0.97
0.0			0.02	0.00	0.08

0.1	**3**	2.27 (1.27-4.43)	0.04	0.00	0.13
0.2			0.07	0.00	0.24
0.3			0.14	0.01	0.45
0.4			0.26	0.01	0.71
0.5			0.40	0.02	0.90
0.6			0.55	0.04	0.97
0.7			0.66	0.06	0.99
0.8			0.75	0.08	1.00
0.9			0.81	0.12	1.00
1.0			0.86	0.16	1.00
Winter					
0.0	**KG**	2.06 (1.38-3.22)	0.01	0.00	0.05
0.1			0.02	0.00	0.07
0.2			0.03	0.00	0.10
0.3			0.04	0.00	0.15
0.4			0.08	0.01	0.20
0.5			0.13	0.04	0.28
0.6			0.23	0.09	0.39
0.7			0.38	0.19	0.56
0.8			0.55	0.29	0.77
0.9			0.70	0.41	0.90
1.0			0.81	0.51	0.97
0.0	**1**	2.03 (1.54-2.82)	0.01	0.00	0.03
0.1			0.01	0.00	0.04
0.2			0.02	0.00	0.06
0.3			0.04	0.01	0.10
0.4			0.06	0.02	0.14
0.5			0.12	0.04	0.21
0.6			0.20	0.10	0.31
0.7			0.33	0.22	0.45
0.8			0.50	0.38	0.62
0.9			0.67	0.52	0.80
1.0			0.80	0.64	0.91
0.0	**2**	1.61 (1.27-2.12)	0.05	0.01	0.16
0.1			0.07	0.01	0.20
0.2			0.11	0.03	0.25
0.3			0.15	0.05	0.30
0.4			0.22	0.09	0.38
0.5			0.31	0.15	0.48
0.6			0.41	0.23	0.61

0.7			0.53	0.31	0.73
0.8			0.64	0.39	0.84
0.9			0.73	0.47	0.91
1.0			0.80	0.54	0.95
0.0	**3**	1.46 (1.14-1.88)	0.08	0.01	0.22
0.1			0.11	0.02	0.25
0.2			0.14	0.04	0.30
0.3			0.19	0.06	0.37
0.4			0.25	0.09	0.46
0.5			0.33	0.13	0.56
0.6			0.41	0.17	0.66
0.7			0.50	0.22	0.76
0.8			0.58	0.26	0.85
0.9			0.66	0.31	0.91
1.0			0.72	0.35	0.95
Spring					
0.0	**KG**	1.92 (1.26-3.17)	0.03	0.00	0.22
0.1			0.04	0.00	0.27
0.2			0.06	0.00	0.32
0.3			0.09	0.00	0.38
0.4			0.14	0.01	0.45
0.5			0.21	0.03	0.53
0.6			0.31	0.08	0.61
0.7			0.44	0.18	0.73
0.8			0.59	0.32	0.84
0.9			0.73	0.47	0.92
1.0			0.82	0.58	0.97
0.0	**1**	1.99 (1.36-3.01)	0.01	0.00	0.05
0.1			0.01	0.00	0.07
0.2			0.02	0.00	0.10
0.3			0.03	0.00	0.13
0.4			0.06	0.01	0.18
0.5			0.09	0.02	0.24
0.6			0.16	0.05	0.32
0.7			0.26	0.11	0.44
0.8			0.40	0.21	0.61
0.9			0.56	0.34	0.79
1.0			0.71	0.46	0.91
0.0	**2**	1.89 (1.40-2.65)	0.01	0.00	0.07
0.1			0.02	0.00	0.10

0.2			0.04	0.00	0.13
0.3			0.06	0.01	0.18
0.4			0.09	0.02	0.25
0.5			0.15	0.04	0.34
0.6			0.25	0.09	0.46
0.7			0.37	0.17	0.60
0.8			0.52	0.29	0.75
0.9			0.66	0.42	0.86
1.0			0.78	0.54	0.94
0.0			0.04	0.00	0.15
0.1			0.06	0.01	0.18
0.2			0.08	0.01	0.20
0.3			0.10	0.02	0.24
0.4			0.13	0.04	0.28
0.5		1.47 (1.15-1.95)	0.18	0.06	0.33
0.6	**3**		0.24	0.10	0.40
0.7			0.31	0.14	0.49
0.8			0.40	0.19	0.61
0.9			0.49	0.24	0.73
1.0			0.57	0.29	0.83

Note. Credible intervals can be interpreted as actual probabilities (i.e., there is a 95% chance the estimate falls between the values of the 2.5 and 97.5 percentiles). Odds ratios reflect the increase in odds of meeting >.75 in spring for a +.1 increase in Mathway score relative to not increasing .1. Odds ratios are constant with respect to increases in Mathway score (increasing from .8 to .9 produces the same odds ratio as increasing from .0 to .1). Odds ratios >1 indicate an increase in the likelihood o meeting the benchmark; odds ratios = 1 indicate equal odds; odds ratios <1 indicate decreases in likelihood. Credible intervals can be interpreted as actual probabilities (i.e., there is a 95% chance that the predicted values fall in the reported intervals). Odds ratio 95% intervals that include a substantial number of values <1 indicate the odds ratio should be interpreted as essentially equal to 1.

Table A3. Predicted Probabilities for Spring 75% by Mathway Score

Mathway Score	Grade	Odds Ratio (95% Interval)	Mean Probability	95% Credible Interval 2.50%	95% Credible Interval 97.5%
Fall					
0.0	**KG**	1.18 (1.01-1.39)	0.66	0.50	0.81
0.1			0.70	0.57	0.82
0.2			0.73	0.62	0.84
0.3			0.77	0.66	0.86
0.4			0.79	0.69	0.88
0.5			0.82	0.71	0.91
0.6			0.84	0.73	0.93
0.7			0.86	0.73	0.95
0.8			0.88	0.74	0.96
0.9			0.89	0.75	0.97
1.0			0.90	0.76	0.98
0.0	**1**	1.18 (0.980-1.43)	0.58	0.38	0.78
0.1			0.62	0.46	0.78
0.2			0.66	0.53	0.79
0.3			0.70	0.59	0.81
0.4			0.73	0.63	0.83
0.5			0.76	0.65	0.85
0.6			0.79	0.67	0.88
0.7			0.81	0.67	0.91
0.8			0.83	0.67	0.93
0.9			0.85	0.67	0.95
1.0			0.86	0.67	0.97
0.0	**2**	1.66 (1.36-2.05)	0.27	0.16	0.42
0.1			0.38	0.26	0.51
0.2			0.50	0.39	0.62
0.3			0.63	0.52	0.73
0.4			0.73	0.63	0.83
0.5			0.82	0.71	0.90
0.6			0.88	0.78	0.95
0.7			0.92	0.83	0.97
0.8			0.95	0.87	0.99
0.9			0.97	0.90	0.99
1.0			0.98	0.93	1.00
0.0			0.32	0.19	0.48

0.1	**3**	1.43 (1.08-1.96)	0.40	0.28	0.54
0.2			0.49	0.36	0.62
0.3			0.58	0.41	0.74
0.4			0.66	0.45	0.84
0.5			0.72	0.48	0.90
0.6			0.78	0.50	0.95
0.7			0.82	0.53	0.97
0.8			0.86	0.55	0.99
0.9			0.88	0.57	0.99
1.0			0.90	0.59	1.00
Winter					
0.0	**KG**	1.42 (1.22-1.66)	0.33	0.15	0.57
0.1			0.41	0.22	0.62
0.2			0.49	0.31	0.68
0.3			0.57	0.41	0.73
0.4			0.66	0.52	0.78
0.5			0.73	0.61	0.83
0.6			0.79	0.69	0.88
0.7			0.84	0.75	0.92
0.8			0.88	0.80	0.94
0.9			0.91	0.84	0.96
1.0			0.94	0.87	0.98
0.0	**1**	1.29 (1.17-1.43)	0.35	0.20	0.52
0.1			0.41	0.27	0.57
0.2			0.47	0.34	0.62
0.3			0.53	0.41	0.66
0.4			0.60	0.49	0.70
0.5			0.66	0.56	0.75
0.6			0.71	0.62	0.80
0.7			0.76	0.67	0.83
0.8			0.80	0.72	0.87
0.9			0.84	0.76	0.90
1.0			0.87	0.79	0.93
0.0	**2**	1.57 (1.38-1.80)	0.11	0.05	0.19
0.1			0.16	0.09	0.25
0.2			0.22	0.14	0.32
0.3			0.31	0.22	0.41
0.4			0.41	0.32	0.51
0.5			0.52	0.43	0.62
0.6			0.63	0.54	0.73

0.7			0.73	0.63	0.82
0.8			0.81	0.71	0.88
0.9			0.87	0.78	0.93
1.0			0.91	0.83	0.96
0.0			0.08	0.04	0.13
0.1			0.12	0.07	0.18
0.2			0.19	0.12	0.26
0.3			0.27	0.19	0.36
0.4			0.39	0.29	0.48
0.5		1.67 (1.49-1.90)	0.51	0.40	0.61
0.6	**3**		0.63	0.52	0.74
0.7			0.74	0.63	0.84
0.8			0.82	0.72	0.90
0.9			0.89	0.79	0.95
1.0			0.93	0.85	0.97

Note. Credible intervals can be interpreted as actual probabilities (i.e., there is a 95% chance the estimate falls between the values of the 2.5 and 97.5 percentiles). Odds ratios reflect the increase in odds of meeting >.75 in spring for a +.1 increase in Mathway score relative to not increasing .1. Odds ratios are constant with respect to increases in Mathway score (increasing from .8 to .9 produces the same odds ratio as increasing from .0 to .1). Odds ratios >1 indicate an increase in the likelihood o meeting the benchmark; odds ratios = 1 indicate equal odds; odds ratios <1 indicate decreases in likelihood. Credible intervals can be interpreted as actual probabilities (i.e., there is a 95% chance that the predicted values fall in the reported intervals). Odds ratio 95% intervals that include a substantial number of values <1 indicate the odds ratio should be interpreted as essentially equal to 1.

Table A4. Grade 4-8 Score Percentiles

Season	Percentile	Mathway Score				
		Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Fall	0.00	0.00	0.00	0.00	0.00	0.00
	0.05	0.03	0.00	0.00	0.00	0.00
	0.10	0.06	0.00	0.02	0.00	0.00
	0.15	0.09	0.03	0.03	0.00	0.00
	0.20	0.12	0.03	0.03	0.03	0.00
	0.25	0.15	0.03	0.07	0.03	0.03
	0.30	0.15	0.03	0.07	0.05	0.03
	0.35	0.18	0.06	0.07	0.07	0.03
	0.40	0.18	0.06	0.09	0.10	0.06
	0.45	0.18	0.06	0.10	0.10	0.06
	0.50	0.18	0.09	0.10	0.14	0.07
	0.55	0.21	0.09	0.10	0.14	0.09
	0.60	0.21	0.12	0.13	0.17	0.09
	0.65	0.21	0.12	0.13	0.17	0.09
	0.70	0.24	0.15	0.17	0.17	0.12
	0.75	0.24	0.15	0.20	0.21	0.15
	0.80	0.26	0.18	0.22	0.21	0.18
	0.85	0.26	0.21	0.23	0.23	0.21
	0.90	0.29	0.26	0.27	0.35	0.29
	0.95	0.32	0.32	0.30	0.50	0.38
	1.00	0.65	0.50	0.50	0.62	0.53
	Percentile	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Winter	0.00	0.00	0.00	0.00	0.00	0.00
	0.05	0.06	0.00	0.00	0.03	0.00
	0.10	0.12	0.03	0.03	0.03	0.03
	0.15	0.15	0.03	0.07	0.07	0.03
	0.20	0.18	0.03	0.10	0.07	0.03
	0.25	0.18	0.03	0.11	0.08	0.06
	0.30	0.21	0.06	0.17	0.10	0.06
	0.35	0.21	0.06	0.17	0.14	0.06
	0.40	0.24	0.09	0.20	0.17	0.09
	0.45	0.26	0.12	0.23	0.17	0.12
	0.50	0.29	0.15	0.23	0.21	0.12
	0.55	0.29	0.18	0.27	0.21	0.15
	0.60	0.32	0.21	0.29	0.24	0.18

	0.65	0.34	0.26	0.30	0.28	0.21
	0.70	0.35	0.29	0.30	0.28	0.24
	0.75	0.38	0.32	0.33	0.34	0.26
	0.80	0.41	0.35	0.37	0.41	0.29
	0.85	0.47	0.38	0.43	0.43	0.32
	0.90	0.50	0.44	0.50	0.46	0.38
	0.95	0.56	0.50	0.63	0.49	0.41
	1.00	0.76	0.88	0.90	0.66	0.56
	Percentile	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**
Spring	0.00	0.00	0.00	0.00	0.00	0.00
	0.05	0.12	0.03	0.00	0.03	0.03
	0.10	0.18	0.06	0.03	0.07	0.03
	0.15	0.24	0.09	0.07	0.07	0.06
	0.20	0.26	0.12	0.12	0.10	0.06
	0.25	0.29	0.12	0.17	0.14	0.07
	0.30	0.34	0.15	0.20	0.17	0.09
	0.35	0.38	0.18	0.23	0.21	0.09
	0.40	0.41	0.21	0.27	0.21	0.09
	0.45	0.44	0.24	0.27	0.24	0.11
	0.50	0.47	0.24	0.30	0.24	0.12
	0.55	0.50	0.26	0.30	0.28	0.16
	0.60	0.53	0.29	0.33	0.31	0.18
	0.65	0.53	0.32	0.37	0.34	0.21
	0.70	0.56	0.32	0.40	0.39	0.21
	0.75	0.59	0.35	0.43	0.44	0.26
	0.80	0.62	0.38	0.47	0.49	0.29
	0.85	0.65	0.41	0.53	0.52	0.35
	0.90	0.68	0.44	0.63	0.59	0.39
	0.95	0.76	0.53	0.73	0.62	0.50
	1.00	0.94	0.76	0.90	0.93	0.76

Table A5. Predicted STAR Benchmark Probabilities by Mathway Score

Mathway Score	Grade	Odds Ratio (95% Interval)	Mean Probability	95% Credible Interval 2.50%	97.5
Fall					
0.0	**4**	2.21 (1.20 – 4.33)	0.00	0.00	0.01
0.1			0.00	0.00	0.01
0.2			0.00	0.00	0.02
0.3			0.01	0.00	0.03
0.4			0.02	0.00	0.08
0.5			0.04	0.00	0.20
0.6			0.09	0.00	0.46
0.7			0.18	0.00	0.75
0.8			0.28	0.00	0.92
0.9			0.39	0.00	0.98
1.0			0.49	0.00	0.99
0.0	**5**	2.92 (1.53-5.77)	0.00	0.00	0.01
0.1			0.00	0.00	0.02
0.2			0.01	0.00	0.03
0.3			0.04	0.00	0.09
0.4			0.10	0.01	0.28
0.5			0.24	0.01	0.65
0.6			0.44	0.03	0.90
0.7			0.62	0.05	0.98
0.8			0.75	0.08	1.00
0.9			0.83	0.12	1.00
1.0			0.89	0.19	1.00
0.0	**6**	1.98 (0.81 – 4.96)	0.01	0.00	0.03
0.1			0.01	0.00	0.04
0.2			0.02	0.00	0.06
0.3			0.04	0.00	0.14
0.4			0.09	0.00	0.34
0.5			0.18	0.00	0.68
0.6			0.29	0.00	0.90
0.7			0.40	0.00	0.98
0.8			0.49	0.00	0.99
0.9			0.56	0.00	1.00
1.0			0.62	0.00	1.00
0.0			0.00	0.00	0.03

0.1	**7**	1.64 (0.63-4.24)	0.00	0.00	0.03
0.2			0.01	0.00	0.03
0.3			0.01	0.00	0.05
0.4			0.02	0.00	0.10
0.5			0.04	0.00	0.21
0.6			0.07	0.00	0.45
0.7			0.13	0.00	0.75
0.8			0.19	0.00	0.91
0.9			0.26	0.00	0.97
1.0			0.33	0.00	0.99
Winter					
0.0	**4**	1.67 (1.43-1.96)	0.03	0.01	0.05
0.1			0.05	0.03	0.08
0.2			0.08	0.05	0.11
0.3			0.12	0.09	0.16
0.4			0.19	0.14	0.24
0.5			0.28	0.21	0.35
0.6			0.40	0.29	0.51
0.7			0.52	0.37	0.66
0.8			0.64	0.46	0.79
0.9			0.74	0.56	0.88
1.0			0.82	0.64	0.93
0.0	**5**	1.83 (1.53-2.21)	0.02	0.01	0.03
0.1			0.03	0.01	0.05
0.2			0.05	0.03	0.08
0.3			0.09	0.05	0.12
0.4			0.15	0.09	0.21
0.5			0.24	0.15	0.34
0.6			0.37	0.23	0.52
0.7			0.51	0.32	0.70
0.8			0.65	0.43	0.83
0.9			0.77	0.54	0.92
1.0			0.85	0.64	0.96
0.0	**6**	1.58 (1.23-2.03)	0.03	0.01	0.08
0.1			0.05	0.02	0.10
0.2			0.08	0.02	0.14
0.3			0.11	0.04	0.20
0.4			0.17	0.05	0.29
0.5			0.24	0.08	0.43
0.6			0.33	0.10	0.58

0.7			0.43	0.13	0.73
0.8			0.53	0.16	0.84
0.9			0.63	0.19	0.91
1.0			0.70	0.23	0.95
0.0			0.01	0.00	0.03
0.1			0.02	0.00	0.05
0.2			0.03	0.01	0.08
0.3			0.07	0.01	0.13
0.4			0.13	0.03	0.24
0.5	**7**	2.11 (1.42-3.29)	0.23	0.06	0.45
0.6			0.38	0.10	0.69
0.7			0.54	0.15	0.87
0.8			0.68	0.21	0.95
0.9			0.79	0.28	0.99
1.0			0.86	0.36	1.00
0.0			0.02	0.00	0.08
0.1			0.04	0.00	0.10
0.2			0.06	0.01	0.14
0.3			0.09	0.01	0.22
0.4			0.15	0.02	0.37
0.5	**8**	1.68 (1.04-2.73)	0.23	0.02	0.57
0.6			0.33	0.02	0.77
0.7			0.43	0.03	0.89
0.8			0.52	0.03	0.95
0.9			0.60	0.03	0.98
1.0			0.67	0.04	0.99
Spring					
0.0			0.02	0.01	0.04
0.1			0.03	0.02	0.05
0.2			0.05	0.03	0.08
0.3			0.09	0.06	0.12
0.4			0.14	0.10	0.18
0.5	**4**	1.74 (1.53-2.00)	0.22	0.17	0.27
0.6			0.33	0.26	0.40
0.7			0.46	0.37	0.55
0.8			0.60	0.48	0.70
0.9			0.72	0.60	0.82
1.0			0.81	0.70	0.90
0.0	**5**	2.26 (1.87-2.80)	0.01	0.00	0.02
0.1			0.02	0.01	0.04

0.2			0.05	0.03	0.08
0.3			0.11	0.07	0.15
0.4			0.22	0.16	0.29
0.5			0.39	0.28	0.50
0.6			0.59	0.44	0.72
0.7			0.76	0.60	0.87
0.8			0.87	0.74	0.95
0.9			0.94	0.85	0.98
1.0			0.97	0.91	0.99
0.0	**6**	1.68 (1.35-2.15)	0.02	0.01	0.06
0.1			0.04	0.01	0.08
0.2			0.06	0.02	0.11
0.3			0.10	0.04	0.16
0.4			0.15	0.07	0.23
0.5			0.23	0.12	0.34
0.6			0.33	0.18	0.49
0.7			0.45	0.24	0.65
0.8			0.57	0.32	0.79
0.9			0.68	0.39	0.89
1.0			0.77	0.47	0.94
0.0	**7**	2.96 (1.67-6.12)	0.00	0.00	0.01
0.1			0.00	0.00	0.01
0.2			0.00	0.00	0.02
0.3			0.01	0.00	0.04
0.4			0.02	0.00	0.07
0.5			0.06	0.01	0.15
0.6			0.15	0.02	0.36
0.7			0.33	0.06	0.69
0.8			0.55	0.13	0.91
0.9			0.73	0.22	0.98
1.0			0.85	0.36	1.00
0.0	**8**	2.00 (1.28-3.25)	0.02	0.00	0.05
0.1			0.03	0.00	0.08
0.2			0.05	0.01	0.12
0.3			0.09	0.02	0.20
0.4			0.16	0.04	0.36
0.5			0.28	0.07	0.58
0.6			0.42	0.10	0.80
0.7			0.57	0.14	0.92
0.8			0.69	0.18	0.97

0.9		0.78	0.22	0.99
1.0		0.84	0.27	1.00

Note: No students met the fall Star benchmark in Grade 8. Credible intervals can be interpreted as actual probabilities (i.e., there is a 95% chance that the predicted values fall in the reported intervals). Odds ratios reflect the increase in odds of meeting >.75 in spring for a +.1 increase in Mathway score relative to not increasing .1. Odds ratios are constant with respect to increases in Mathway score (increasing from .8 to .9 produces the same odds ratio as increasing from .0 to .1). Odds ratios >1 indicate an increase in the likelihood o meeting the benchmark; odds ratios = 1 indicate equal odds; odds ratios <1 indicate decreases in likelihood. Credible intervals can be interpreted as actual probabilities (i.e., there is a 95% chance that the predicted values fall in the reported intervals). Odds ratio 95% intervals that include a substantial number of values <1 indicate the odds ratio should be interpreted as essentially equal to 1.

Table A6. Predicted Spring 75% Probabilities by Mathway Score

Mathway Score	Grade	Odds Ratio (95% Interval	Mean Probability	95% Credible Interval 2.50%	95% Credible Interval 97.5
Fall					
0.0	**4**	2.09 (0.91-5.13)	0.01	0.00	0.06
0.1			0.02	0.00	0.06
0.2			0.03	0.00	0.08
0.3			0.07	0.01	0.18
0.4			0.14	0.01	0.44
0.5			0.26	0.01	0.78
0.6			0.39	0.01	0.94
0.7			0.51	0.01	0.99
0.8			0.60	0.01	1.00
0.9			0.67	0.01	1.00
1.0			0.72	0.01	1.00
0.0	**6**	0.916 (0.32-2.61)	0.10	0.00	0.32
0.1			0.08	0.00	0.22
0.2			0.08	0.00	0.24
0.3			0.09	0.00	0.36
0.4			0.11	0.00	0.56
0.5			0.13	0.00	0.75
0.6			0.16	0.00	0.87
0.7			0.18	0.00	0.94
0.8			0.21	0.00	0.98
0.9			0.23	0.00	0.99
1.0			0.24	0.00	1.00
0.0	**7**	1.85 (0.84-4.64)	0.01	0.00	0.07
0.1			0.02	0.00	0.08
0.2			0.03	0.00	0.10
0.3			0.04	0.00	0.15
0.4			0.08	0.00	0.29
0.5			0.15	0.00	0.53
0.6			0.24	0.01	0.79
0.7			0.35	0.01	0.93
0.8			0.45	0.00	0.98
0.9			0.53	0.00	1.00
1.0			0.60	0.00	1.00
0.0			0.02	0.00	0.11

0.1	**8**	1.98 (0.59-6.74)	0.02	0.00	0.10
0.2			0.03	0.00	0.14
0.3			0.05	0.00	0.23
0.4			0.09	0.00	0.45
0.5			0.17	0.00	0.73
0.6			0.28	0.00	0.92
0.7			0.39	0.00	0.98
0.8			0.47	0.00	1.00
0.9			0.54	0.00	1.00
1.0			0.59	0.00	1.00
Winter					
0.0	**4**	3.15 (2.02-5.12)	0.00	0.00	0.00
0.1			0.00	0.00	0.01
0.2			0.00	0.00	0.01
0.3			0.01	0.00	0.03
0.4			0.03	0.01	0.06
0.5			0.08	0.03	0.16
0.6			0.22	0.08	0.41
0.7			0.46	0.18	0.75
0.8			0.70	0.33	0.93
0.9			0.85	0.52	0.99
1.0			0.93	0.70	1.00
0.0	**5**	2.88 (1.29-6.99)	0.00	0.00	0.00
0.1			0.00	0.00	0.00
0.2			0.00	0.00	0.00
0.3			0.00	0.00	0.01
0.4			0.00	0.00	0.02
0.5			0.01	0.00	0.04
0.6			0.04	0.00	0.13
0.7			0.11	0.00	0.41
0.8			0.25	0.00	0.80
0.9			0.42	0.00	0.96
1.0			0.57	0.01	0.99
0.0	**6**	1.38 (0.82-2.27)	0.01	0.00	0.03
0.1			0.01	0.00	0.03
0.2			0.01	0.00	0.04
0.3			0.02	0.00	0.06
0.4			0.02	0.00	0.09
0.5			0.04	0.00	0.15
0.6			0.05	0.00	0.26

0.7			0.08	0.00	0.41
0.8			0.12	0.00	0.58
0.9			0.16	0.00	0.74
1.0			0.21	0.00	0.86
0.0	**7**	1.60 (0.657-4.25)	0.01	0.00	0.04
0.1			0.01	0.00	0.03
0.2			0.01	0.00	0.03
0.3			0.01	0.00	0.05
0.4			0.02	0.00	0.08
0.5			0.04	0.00	0.17
0.6			0.08	0.00	0.38
0.7			0.13	0.00	0.68
0.8			0.21	0.00	0.88
0.9			0.28	0.00	0.97
1.0			0.35	0.00	0.99
0.0	**8**	2.81 (0.91-9.91)	0.00	0.00	0.02
0.1			0.00	0.00	0.02
0.2			0.01	0.00	0.03
0.3			0.01	0.00	0.06
0.4			0.03	0.00	0.14
0.5			0.09	0.00	0.39
0.6			0.21	0.00	0.78
0.7			0.37	0.00	0.96
0.8			0.51	0.00	1.00
0.9			0.61	0.00	1.00
1.0			0.69	0.00	1.00

Note. Of those with available data, no students in Fall of Grade 5 also scored above 75% in Spring. Odds ratios reflect the increase in odds of meeting >.75 in spring for a +.1 increase in Mathway score relative to not increasing .1. Odds ratios are constant with respect to increases in Mathway score (increasing from .8 to .9 produces the same odds ratio as increasing from .0 to .1). Odds ratios >1 indicate an increase in the likelihood o meeting the benchmark; odds ratios = 1 indicate equal odds; odds ratios <1 indicate decreases in likelihood. Credible intervals can be interpreted as actual probabilities (i.e., there is a 95% chance that the predicted values fall in the reported intervals). Odds ratio 95% intervals that include a substantial number of values <1 indicate the odds ratio should be interpreted as essentially equal to 1.

Figure A1. Fall Item Proportions Grades K-3

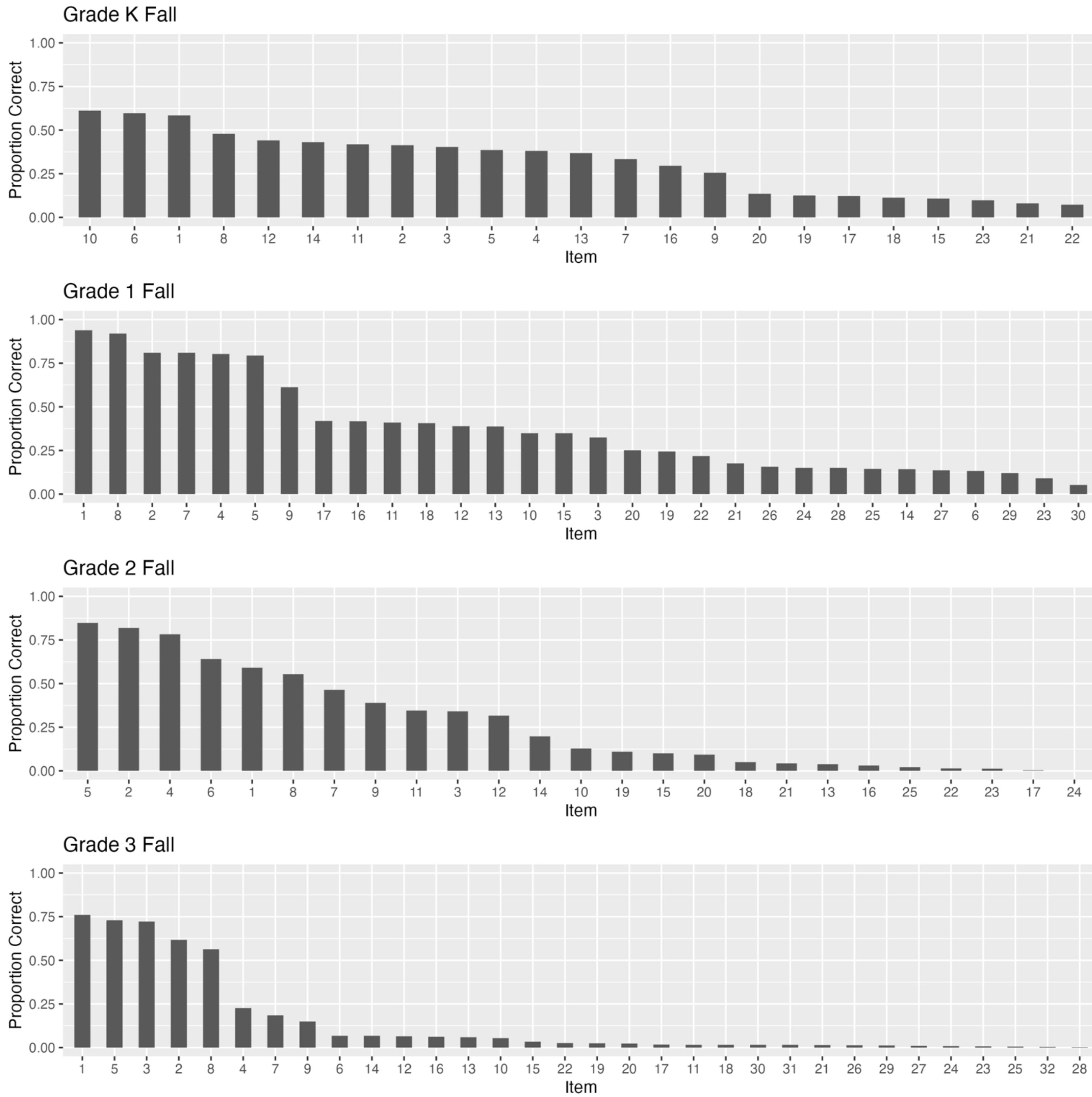


Figure A2. Winter Item Proportions Grades K-3

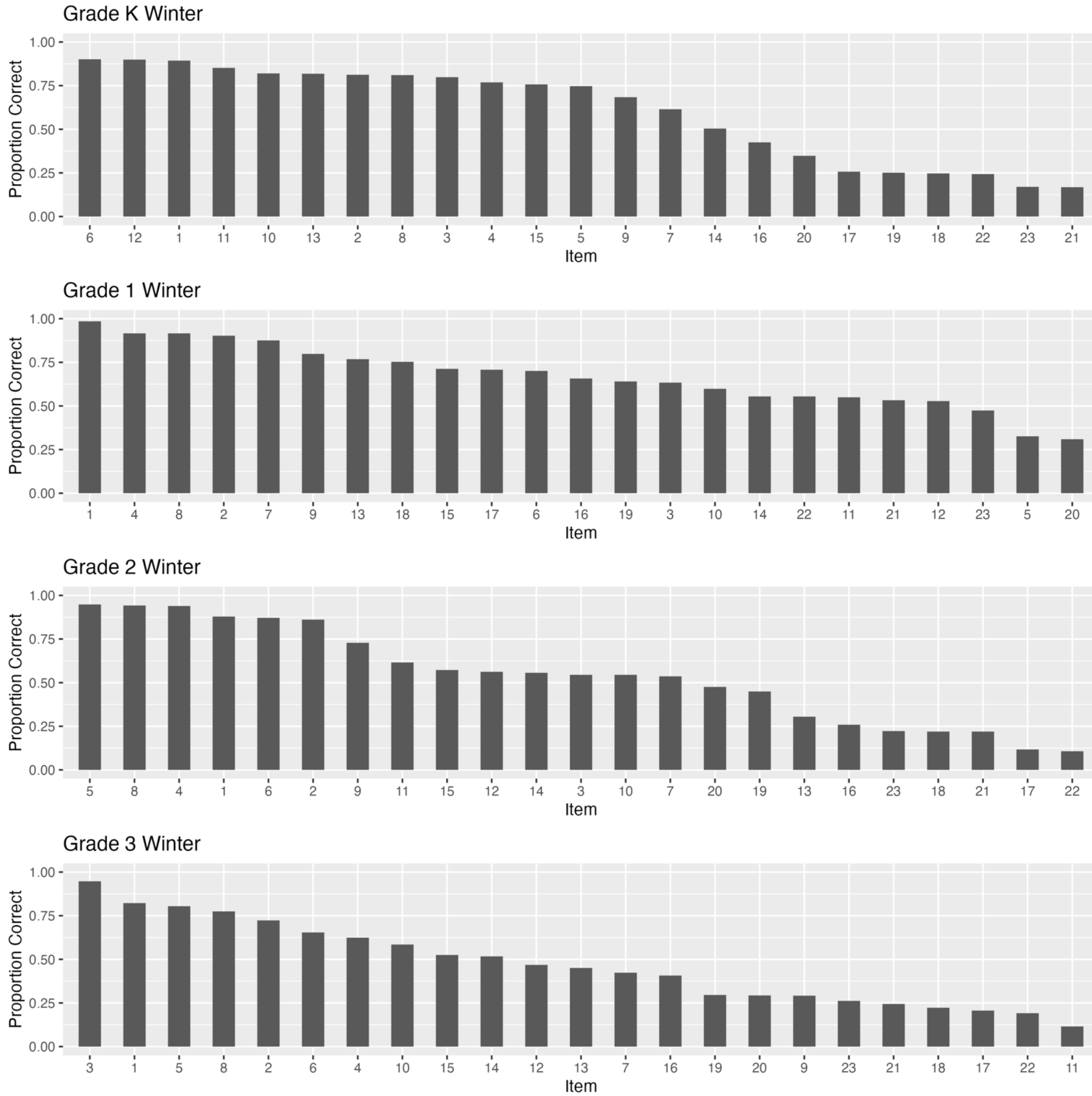


Figure A3. Winter Item Proportions Grades K-3

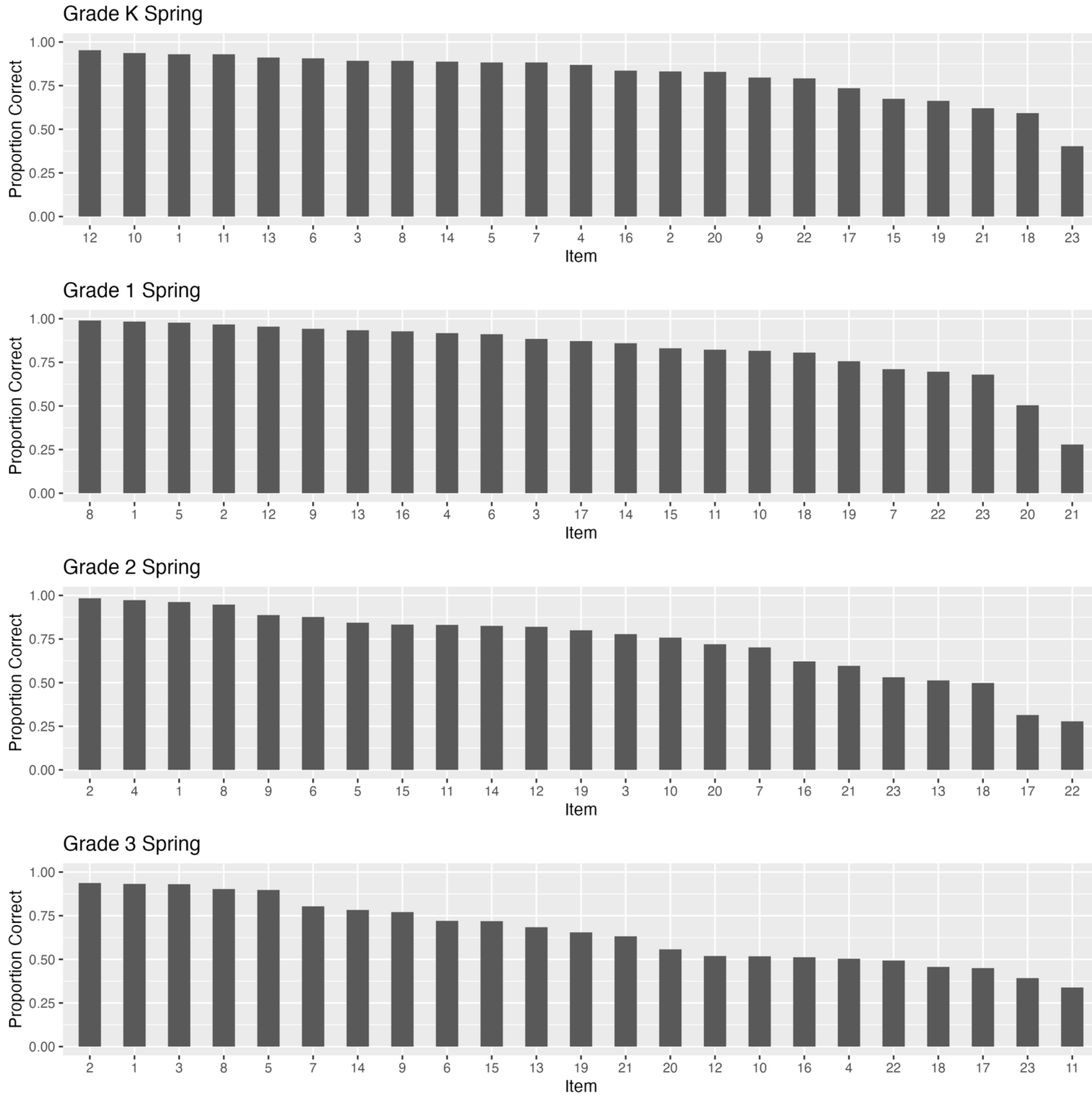


Figure A4. Fall Item Proportions Grades 4-8

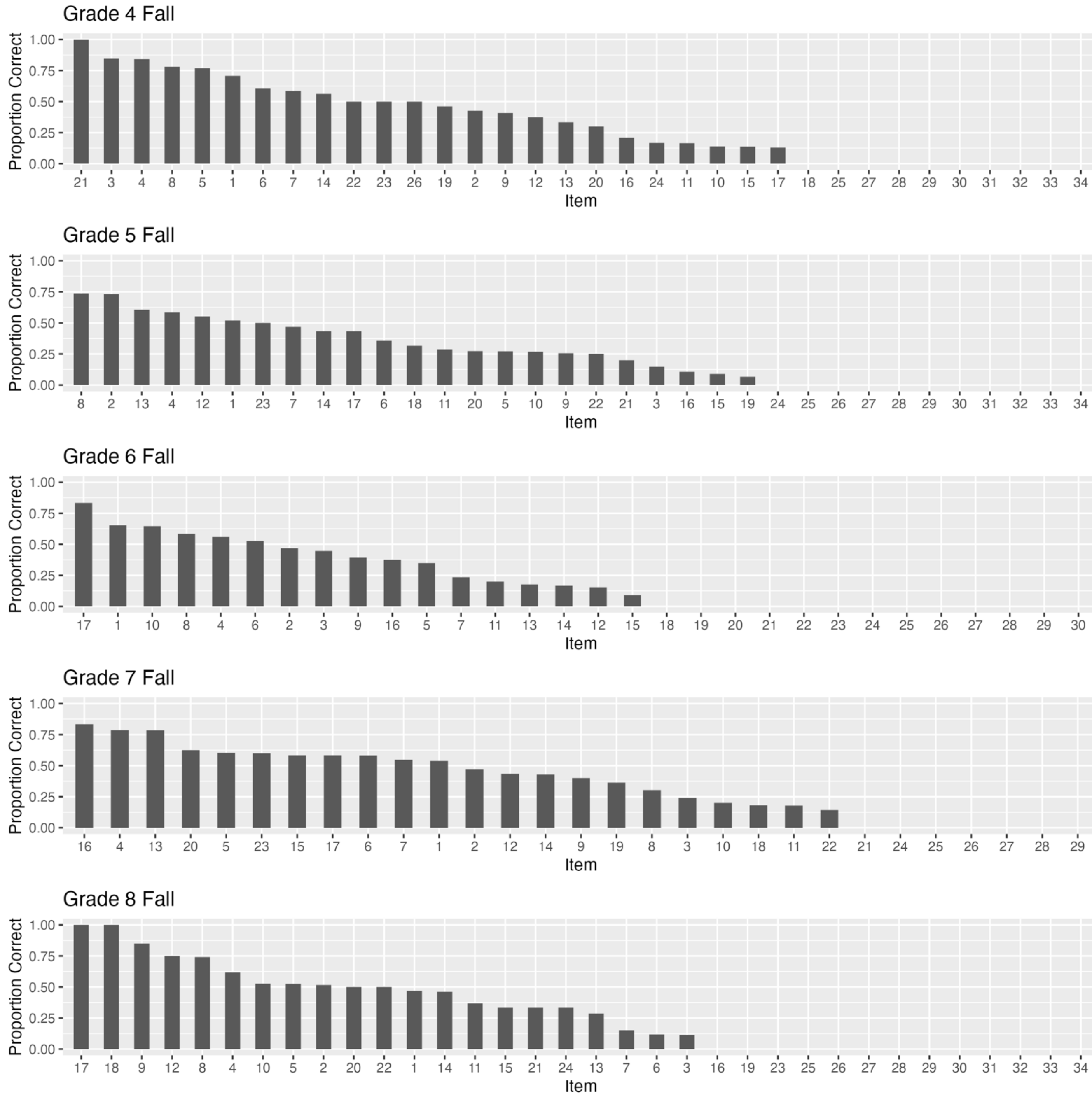


Figure A5. Winter Item Responses Grades 4-8

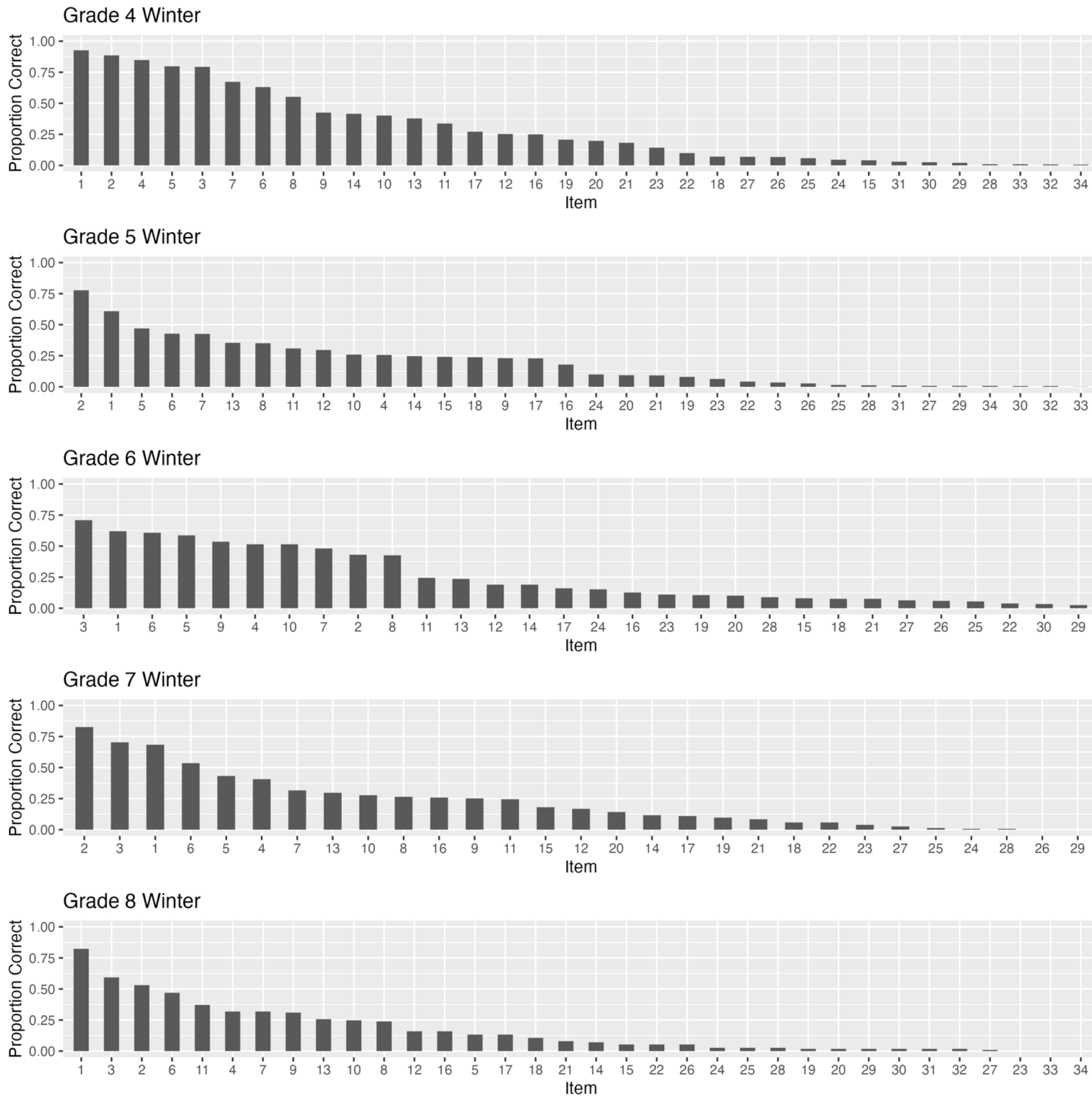


Figure A6. Spring Item Responses Grades 4-8