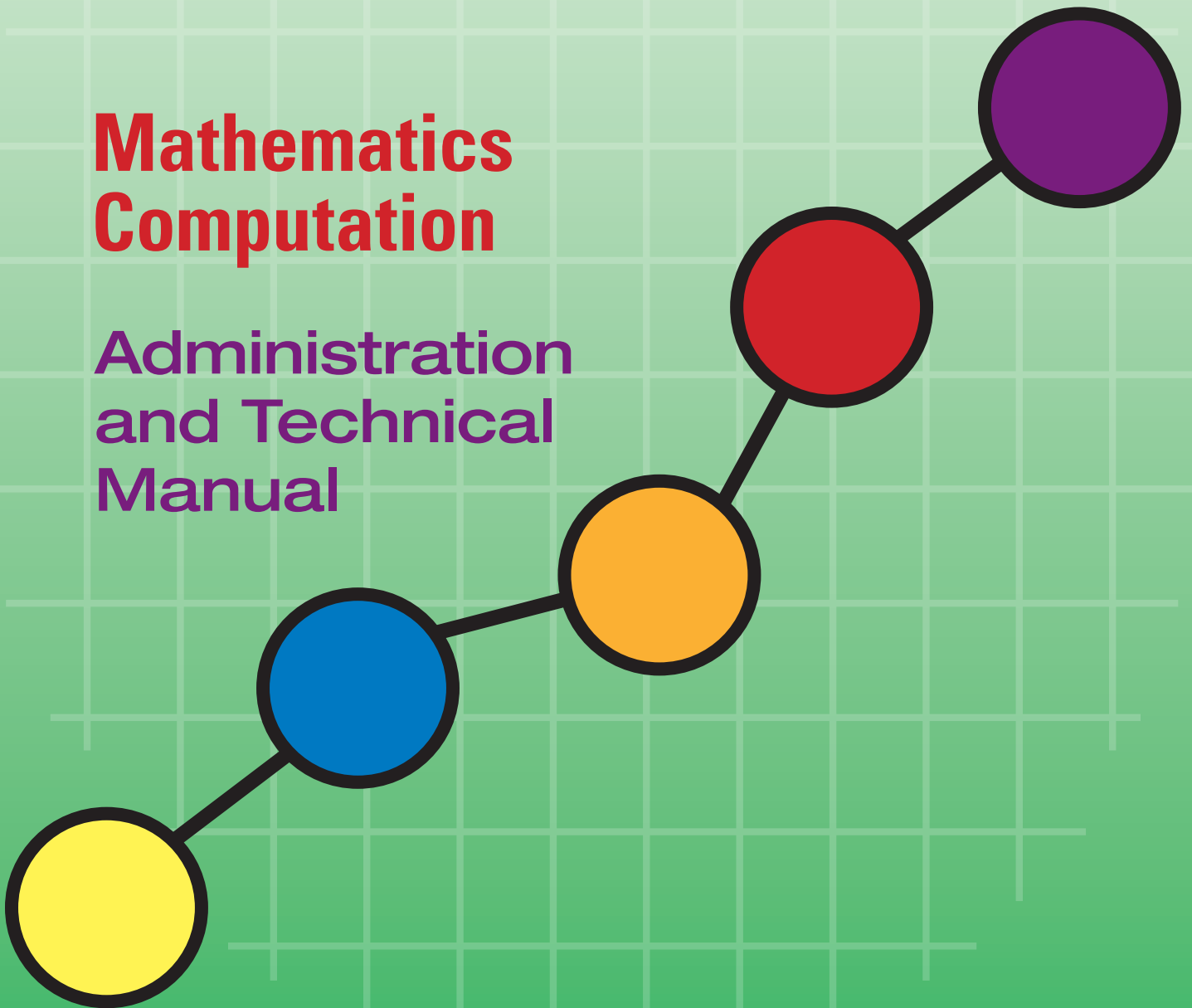




# AIMSweb<sup>®</sup>

## Mathematics Computation

Administration  
and Technical  
Manual



**Pearson Corporate Executive Office 5601 Green Valley Drive Bloomington, MN 55427**  
**800.627.7271 www.PsychCorp.com**

Copyright © 2010 NCS Pearson, Inc. All rights reserved.

**Warning:** No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the copyright owner.

**Pearson**, the **PSI logo**, **PsychCorp**, and **AIMSweb** are trademarks in the U.S. and/or other countries of Pearson Education, Inc., or its affiliate(s).

Portions of this work were previously published.

Printed in the United States of America.

2 3 4 5 6 7 8 9 10 11 12 B C D E

**PEARSON**  




# Table of Contents

## Section 1

<b>Introduction</b>	1
<b>Welcome to AIMSweb®.</b>	1
Mathematics Computation (M-COMP)	1
Changes From the Previous Edition	2
Test Security and Copyright Restrictions	2
User Qualifications	3
Contact Information	3
Accessing More AIMSweb Material	3

## Section 2

<b>Using AIMSweb® Mathematics Computation (M-COMP).</b>	5
Benchmarking and Screening	6
Progress Monitoring	8
Goal Setting	9

## Section 3

<b>Guidelines for Administration, Scoring, and Reporting</b>	11
<b>General Testing Considerations</b>	11
Testing Students With Special Accommodations	11
Conducting a Survey-Level Assessment or Off-Level Testing	12
<b>Administering the Probes</b>	13
Administration Directions	14
<b>Scoring Guidelines</b>	15
How to Score the M-COMP	16
M-COMP Scoring Examples	17
<b>Reporting</b>	26
Generating Student Reports	26

## Section 4


<b>Developing and Standardizing the Test</b>	<b>29</b>
<b>Creating the New Blueprint</b>	<b>29</b>
Item Development	29
Item Review	30
Test Construction	30
<b>Major Research Stages</b>	<b>31</b>
Pilot Studies	31
National Field Test	31
Finalizing and Selecting Probes	32
<b>Appendix A Technical Adequacy and Data Tables</b>	<b>33</b>
A.1 Demographic Characteristics of the Sample by Grade and Geographic Region	34
A.2 Demographic Characteristics of the Sample by Grade and Community Type	34
A.3 Demographic Characteristics of the Sample by Grade and Sex	35
A.4 Demographic Characteristics of the Sample by Grade and Race/Ethnicity	35
A.5 Demographic Characteristics of the Sample by Grade and Median Family Income	36
A.6 Descriptive and Reliability Statistics by Grade	36
A.7 Correlations of AIMSweb M-COMP Scores With Group Mathematics Assessment and Diagnostic Evaluation (G-MADE) Scores by Grade	37
<b>Glossary</b>	<b>39</b>
<b>References</b>	<b>41</b>
<b>Figures</b>	
2.1 Benchmark/Progress Monitor Flow Chart	5
2.2 M-COMP Individual Student Report	7
2.3 M-COMP Progress Monitoring Improvement Report	8
3.1 Scored Answer Key	16
3.2 Correct Answer Not on the Answer Key	17
3.3 Range of Acceptable Answers	18
3.4 Overcorrection Into Error	19
3.5 Targeted Answer in Directions	20
3.6 Mixed Number for Grade 8, Item 22	20
3.7 Answers to Decimal Problems	21
3.8 Written Answers	22
3.9 Crossed-Out Answer	23

3.10 Difficult-to-Read Response . . . . .	24
3.11 Reversed Numbers With Intended Number Obvious . . . . .	25
3.12 Rotated Numbers With Intended Number Indeterminable. . . . .	25

## Tables

3.1 Total Point Value by Grade . . . . .	15
4.1 Domains Evaluated With M-COMP by Grade . . . . .	29
4.2 National Field Testing Item and Probe Count by Grade . . . . .	32





# Section 1

## Introduction

### Welcome to AIMSweb®

---

AIMSweb® is an assessment, data organization, and reporting system that provides the framework and data necessary for response to intervention (RTI) and multitiered instruction. Designed specifically to benchmark and monitor progress, AIMSweb uses Curriculum-Based Measurement (CBM) practices: brief, reliable, and valid measures of basic reading skills, language arts, and mathematics. These standardized tests are based on general outcome measurement principles so they can be efficiently and accurately used to evaluate student progress relative to a year-end target, regardless of curriculum or intervention.

Schools use AIMSweb benchmark assessments to universally screen all students three times a year in order to identify at-risk students early. This enables the educators to align students to multitiered interventions according to each student's instructional need—before he or she has a chance to fail. For students receiving intervention, schools use AIMSweb to monitor their progress frequently, to ensure these students are progressing at the rate necessary to make adequate progress and meet their goal.

AIMSweb is designed to be generalized to any curriculum and is used to assess skills taught across a range of subject matter. As general outcome measures, they serve as *indicators* of performance (e.g., general mathematics skills), as opposed to a mastery test that reports performance level on specific skills or concepts (e.g., division of decimals) or a diagnostic assessment that analyzes performance in depth.

AIMSweb assessment information can be collected through web-based data capture (for some measures) or by individually entering a student's scores (for all measures). Reports tied to robust graphs and tables are generated instantly, can be interpreted by norms-based or standards-based comparisons, and can be viewed anytime the user is connected to the Internet. Individual student data and reports are transportable across time, across schools and communities, and across the kinds of decisions educators typically make about student performance.

### Mathematics Computation (M-COMP)

AIMSweb Mathematics Computation (M-COMP) is a series of assessments that yield general math computation performance and rate of progress information. M-COMP includes three probes for benchmarking and 30 probes for progress monitoring for each grade, 1 through 8. M-COMP is a timed, 8-minute, open-ended, paper-based test that can be group administered or individually administered.

M-COMP is a revision of the AIMSweb Mathematics-CBM (M-CBM) and Mathematics-CBM2 (M-CBM2), which will be retired in fall of 2011. Features of M-COMP include:

- Enhanced content to increase depth of information and to align more closely with National Council of Teachers of Mathematics (NCTM) standards,
- Updated and streamlined scoring rules,
- Clearer, more student-friendly format, and
- Scores equated to M-CBM and M-CBM2.

## Changes From the Previous Edition

Feedback from M-CBM and the M-CBM2 users was considered when designing this revision. Users noted concerns in (1) the length of time and difficulty of scoring, (2) students skipping around to easier items skewing performance data, and (3) the format of the test impeded student performance and the scoring process. These three major concerns were taken into account and addressed during the conceptual development of the M-COMP.

The concern regarding the time-intensive and potentially subjective nature of counting every numeral in Correct Digit scoring and the Critical Process scoring was taken into account when the team developed AIMSweb Mathematic Concepts and Applications (M-CAP, 2009), and a more streamlined scoring system was developed for that test. Because user reaction has been so favorable to the M-CAP scoring, the AIMSweb content team evaluated its efficacy for use in the development of M-COMP.

Based on user feedback for M-CBM and the M-CBM2 and positive user reaction to the M-CAP, the AIMSweb content team re-evaluated the current M-CBM and M-CBM2 scoring process against the weighted-scoring system as used in M-CAP, and the data indicated that when applied to the M-COMP, the system minimized scoring time, maximized sensitivity to growth, controlled for students who skip to the easiest items, and ensures the psychometric soundness of the process. Applying this system therefore addressed the first two concerns noted by current M-CBM and M-CBM2 users.

The third concern noted was that the M-CBM and M-CBM2 test format impeded the students' ability to take the test easily and the teachers' ability to score the test quickly. To address these concerns, the format was revised by placing boxes around each item, numbering the items, and increasing the space between items.

## Test Security and Copyright Restrictions

Educators who use the M-COMP are responsible for ensuring that the test materials, including the probes, remain secure. Individual or collective test items may not be copied or disclosed to anyone but the appropriate school personnel; to do so would compromise the validity of the M-COMP measure and reduce its value as a measurement tool. Under no circumstances may test materials be resold or displayed in locations where unqualified individuals can purchase or view partial or complete portions of the M-COMP. This includes personal websites and Internet auction sites.



All test items, data, and materials are copyrighted. With the exception of usage as described in this Manual, the Pearson Education, Inc., Rights and Permissions Department must approve in writing (hard copy or electronic communication) the copying or reproduction of any test materials. The only other exception to this requirement is the copying of a completed M–COMP probe(s) to convey a student’s records to a qualified educator.

## User Qualifications

The Mathematics Computation (M–COMP) is designed to be administered by general and special education teachers, but may be used by school psychologists and other education specialists as well. Users must adhere to the standardized administration and scoring rules to obtain accurate results and properly interpret those results. Any deviation from the standardized administration and scoring procedures may invalidate results and the resulting scores cannot be reported or compared to the normative data. Before administering an M–COMP probe, please read Section 3 of this manual in its entirety. Section 3 also provides further guidance in using M–COMP with populations that require special accommodations.

## Contact Information

If you have any questions, comments, or suggestions about the AIMSweb M–COMP or concerns about the policies and procedures presented in this Manual, please contact:

AIMSweb Customer Relations  
Pearson  
19500 Bulverde Road  
San Antonio, TX 78259  
Phone: 866-313-6194  
Fax : 866-313-6197  
Email: [aimswebsupport@pearson.com](mailto:aimswebsupport@pearson.com)

## Accessing More AIMSweb Material

The AIMSweb documents and manuals referenced within this document can be found by logging in to your AIMSweb account, select the yellow **Downloads** tab from the top of the page, and scroll down until you find the document you are looking for. Clicking on the document title will bring up a printable Portable Document File (PDF) of that document.

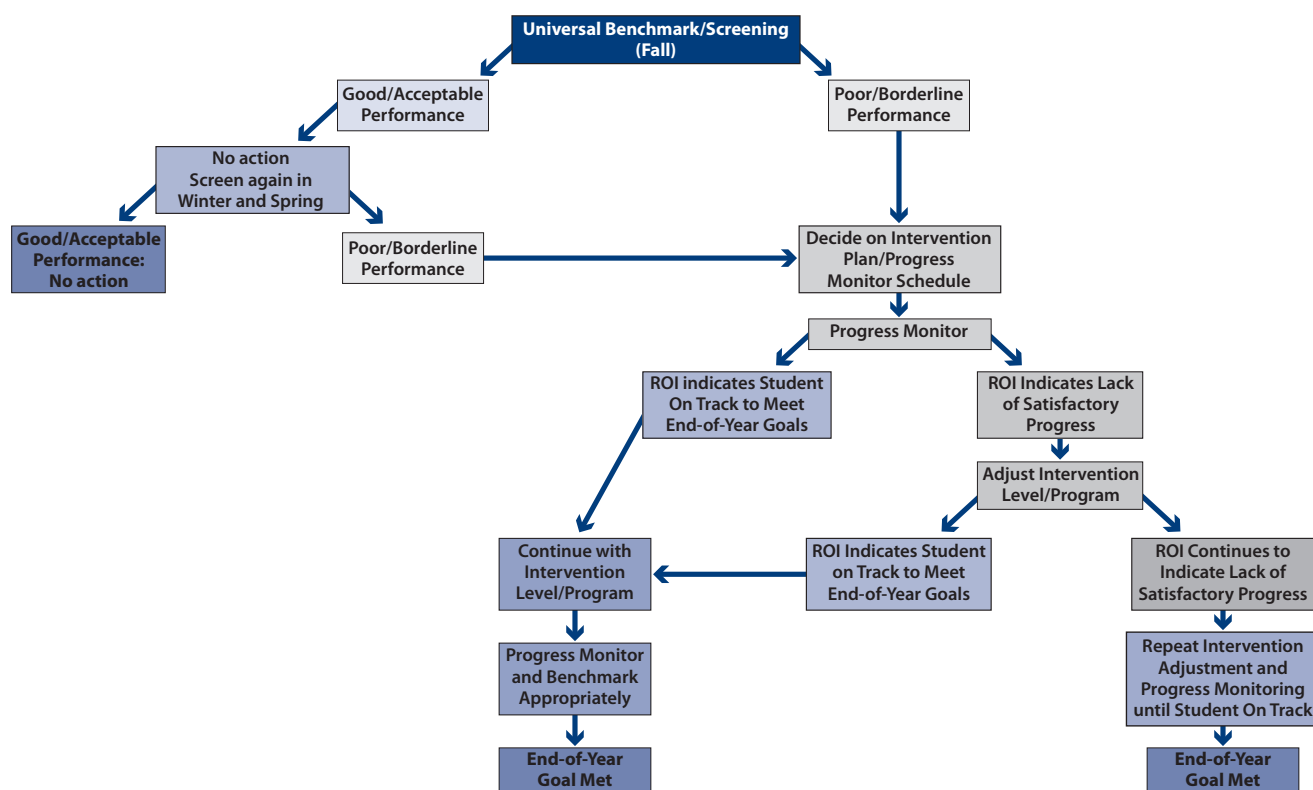


## Section 2

# Using AIMSweb® Mathematics Computation (M-COMP)

Benchmarking students' performance three times a year yields distinct data points, enabling you to determine if students are on track with their progress, are struggling and may benefit from intervention, or are out performing their peers. Progress monitoring for students identified during Benchmarking as likely to benefit from some level of intervention enables you to determine whether or not students are benefiting by measuring their rate of improvement.

Using the AIMSweb reporting system, a desirable rate of improvement (ROI) can be determined and progress monitoring decisions made based on achieving that goal. The ROI is represented by a trend line, or slope, which indicates the average weekly improvement. If the trend line (i.e., the actual ROI) meets or exceeds the aim line (i.e., the expected ROI), you can then be confident that the student is benefitting from the intervention, and that you should continue with the program. If the trend line is not meeting or exceeding the aim line, then you should consider changing the intervention approach (e.g., same program at higher intensity or a different program). See Figure 2.1.



**Figure 2.1 Benchmark/Progress Monitor Flow Chart**

You or another education specialist may determine the frequency of progress monitoring based on the specific needs of a particular child. Progress monitoring can be as frequent as once a week, once a month, or several times between Benchmark periods. For students who are receiving Tier 2 and Tier 3 interventions, the goal material typically is their grade-level material; that is, grade 4 students would be progress monitored using grade 4 M–COMP probes. However, for students who have IEPs with significant mathematics computation discrepancies, goal material may be lower than their current grade placement. In these instances, you should conduct a Survey Level Assessment (SLA) to determine a student's present level of performance. Teams then write an individualized goal to represent the grade-level proficiency to be achieved by the IEP annual review date, which may include progress monitoring with off-grade level probes. See the sections “Benchmarking and Screening” and “Progress Monitoring” that appear later in this section.

## Benchmarking and Screening

---

AIMSweb M–COMP has three Benchmark probes per grade to be administered to all students during the standard school year:

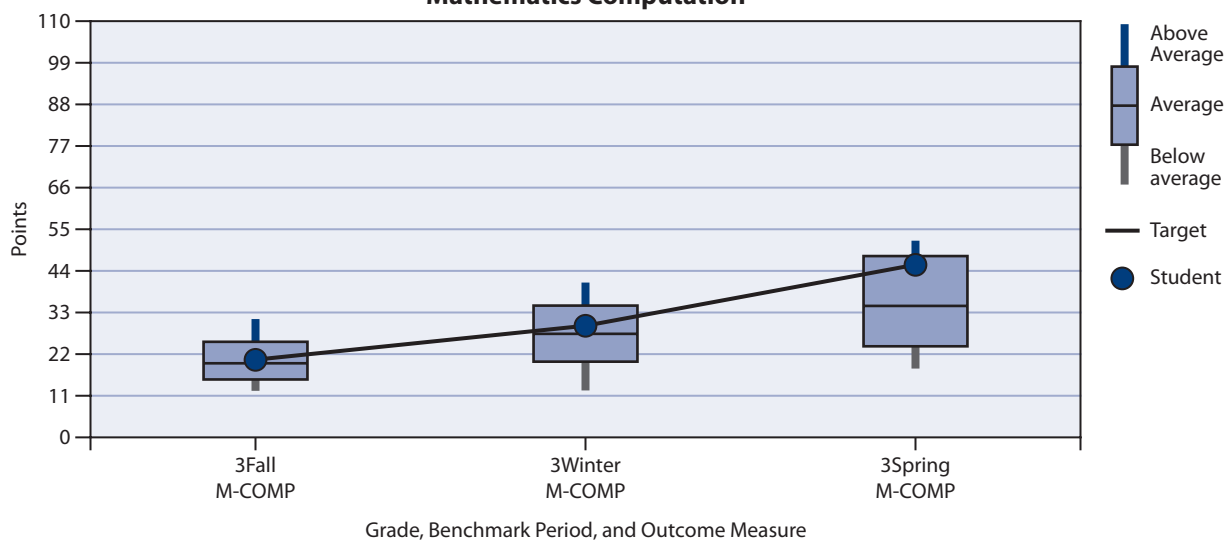
- Fall (September 1–October 15), which may also be used for purposes of screening
- Winter (January 1–January 31), for progress monitoring and program evaluation
- Spring (May 1–May 31), for progress monitoring and program evaluation

The purpose of benchmarking is to ensure that all students are assessed after a similar exposure to the school curriculum. Although benchmarking periods range from 4 to 6 weeks, the process should be completed within 2 weeks after a school begins the benchmarking process. (Most schools complete benchmark testing within 2 days.) This limits the effects of additional instruction for some students. If your school district's schedule differs from the standard schedule, adjust your benchmarking periods to reflect the level of instruction consistent with that of the standard school year and the suggested Benchmark testing dates.

You can use the initial M–COMP Benchmark probe as a screening tool to make RTI decisions, and then compare the results to normative- or standards-based data. Using the normative-based data, an individual student report that presents the range of average M–COMP student performance (i.e., scores between the 25th and 74th percentiles) can be generated that shows a student's current M–COMP performance (the number of points earned on a particular probe). The student's performance can be judged using percentile ranks and the raw scores relative to the normative group, and can be used to make screening decisions.

Figure 2.2 presents a sample student's M–COMP performance relative to a school norm is shown on the standard AIMSweb “box and whisker plot” (note that any individual student's ROI can also be compared to a normative group).

**Teacher:** Teacher One **Student:** Student One  
**Benchmark Scores for 2009–2010 School Year**  
**State School District – Elementary School**  
**Student One (Grade 3)**  
**Compared To: Elementary School**  
**Mathematics Computation**



Copyright © 2009 by NCS Pearson, Inc.

**Benchmark Comparison:** Elementary School

Outcome Measure	Year	Grade	Fall	Winter	Spring	Level of Skill	Instructional Recommendation
Mathematics Computation (M-COMP)	2009–2010	3	20.0	29.0	45.0	Average	Continue Current Program (Elementary School Spring Percentiles)

Student One improved from **20** Points (pts) from Grade 3 Probes at the Fall Benchmark to **45** Points (pts) at the Spring Benchmark. The rate of improvement (ROI) from the Fall Benchmark is **0.7** Points per week. Currently, Student One's score is **Average** compared to Elementary School Spring Percentiles. This was a score **at the 64** percentile compared to other students in the Elementary School.

## Figure 2.2 M-COMP Individual Student Report

This report reflects the student's M-COMP performance level over time, and the number of points earned on a particular probe from fall to winter (or from fall to winter to spring), representing this student's ROI. The line extending from the top of the box represents the range of above average M-COMP student performance (i.e., scores at the 75th percentile and above). Scores above this top line represent scores in the upper 10% of students in the comparison group. The line extending from the bottom of the box represents the range of below average M-COMP student performance (i.e., scores between the 10th and 24th percentiles). Often, scores at this level are used to identify potential candidates for tiered interventions (e.g., Tier 2). Scores below this bottom line represent scores in the lower 10% of students in the comparison group. Typically, scores at this level are used to identify potential candidates for the most intensive of tiered interventions (e.g., Tier 3).

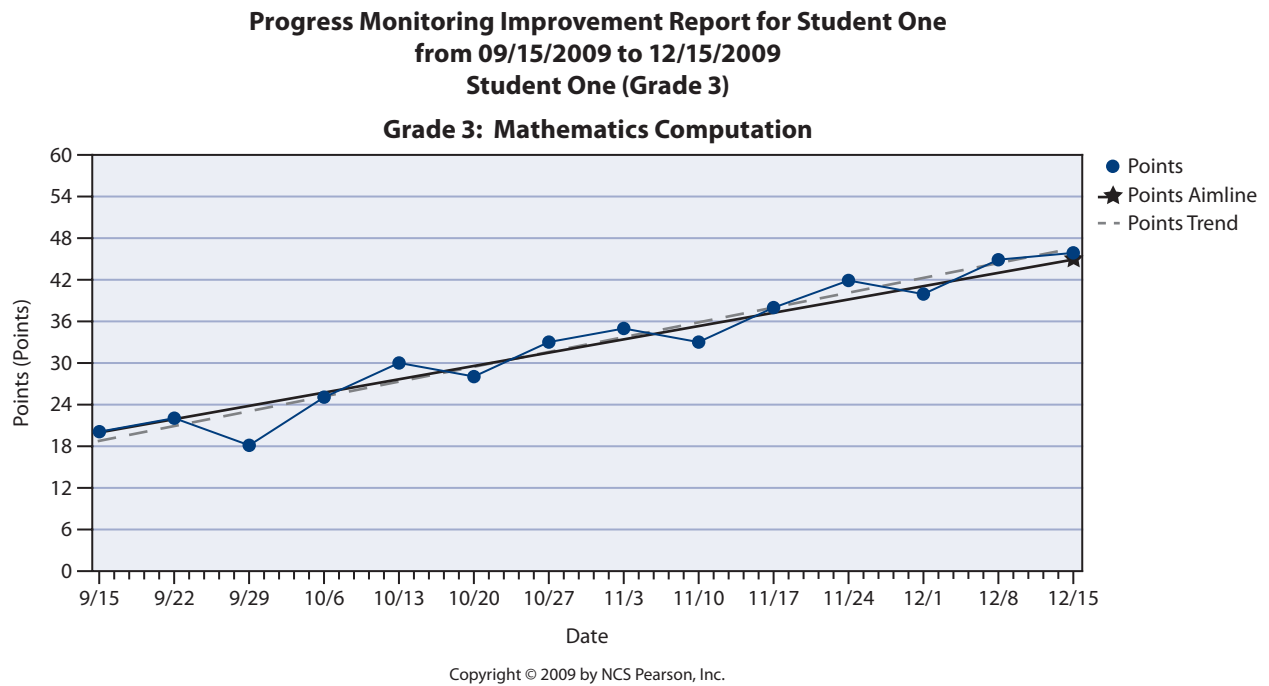
With a standards-based approach to interpretation, educators can use a student's M-COMP Benchmark scores to broadly predict performance on a high-stakes test (e.g., a state-required achievement test) by identifying those students who are most likely to pass and those who are most likely to not pass your state test. More importantly, however, M-COMP Benchmarking better enables you to identify students who are in between the extremes of a performance range, so that students who otherwise may slip through the cracks can get the intervention they need to better equip them for success in class, as well as on state tests.

For more detail on how to use M-COMP in a benchmark assessment approach for screening all students, see the AIMSweb Training Workbook *Organizing and Implementing a Benchmark Assessment Program* (Shinn, 2002a). This and other manuals can be found on the **Downloads** tab within your AIMSweb account.

Progress Monitoring

M-COMP provides 30 Progress Monitoring probes for each grade (Grades 1–8). Like the three Benchmark probes, the Progress Monitoring probes have been standardized to be equivalent in difficulty. When benchmarking and progress monitoring are used together, you can be confident that improvement or lack of improvement in a student’s performance has been accurately tracked.

When a student is identified as potentially at risk and requiring mathematics intervention, you and other qualified education professionals may meet to discuss appropriate end-of-year goals for that student, and determine the required ROI to take the student from his or her current performance level to the desired end-of-year performance level. Figure 2.3 presents a frequent progress monitoring graph that shows an example of how a student’s rate of improvement has been tracked with progress monitoring.



Goal Statement

In **13.0** weeks, Student One will achieve **45** Points from grade **3** Mathematics Computation. The rate of improvement should be **1.92** Points per week. The current average rate of improvement is **2.14** Points per week.

Date	09/15	09/22	09/29	10/06	10/13	10/20	10/27	11/03	11/10	11/17	11/24	12/01	12/08	12/15
Points	20	22	18	25	30	28	33	35	33	38	42	40	45	46
Goal/Trend ROI	1.92/2.14													

Figure 2.3 M-COMP Progress Monitoring Improvement Report

In progress monitoring, you use a single, different M–COMP probe each time. Progress monitoring frequency should be consistent with the intensity level of intervention for that student. You may want to monitor the progress of a student who is in the Tier 2 intervention range once or twice a month. For a student in the more intensive Tier 3 range, you may want to monitor progress once or twice a week.

## Goal Setting

---

To get the most value from progress monitoring, it is important to set meaningful goals. The components of the goals are (1) an established time frame, (2) the level of performance expected, and (3) the criterion for success. Typical time frames include the duration of the intervention or the end of the school year. An annual time frame is typically used when IEP goals are written for students who are receiving special education.

M–COMP goals are written in an individualized yet standard format such as the example below:

In 34 weeks (1 academic year), Betsy will write correct answers to computation problems, earning 40 points on grade 5 M–COMP probes.

You can establish the criterion for success according to standards, local norms, national norms, or a normative rate of improvement. The team may want to compare a student's performance to district/local norms; that is, to compare the scores to his or her peers in the context of daily learning. The last type of criterion is to use a normative rate-of-improvement (ROI). Using a mathematical formula, an average rate of weekly improvement attained from a normative database is multiplied by the time frame to determine the criterion for success.

National norms for the M–COMP will be available in the fall of 2011. For detailed information and direction for setting goals, see *Progress Monitoring Strategies for Writing Individual Goals in General Curriculum and More Frequent Formative Evaluation* (Shinn, 2002b).





## Section

# 3

# Guidelines for Administration, Scoring, and Reporting

## General Testing Considerations

M-COMP is a standardized test; that is, it is an assessment with directions, time limits, materials, and scoring procedures designed to remain consistent each time the test is given, to ensure comparability of scores. To make valid normative decisions (national, state, district, or school), you must follow all directions carefully. Changes in the presentation of the probes, alterations of the instructions given to students, or the inappropriate use of probes as teaching instruments may invalidate any decisions or conclusions about student performance. M-COMP can be administered to whole classes, small groups (3–4 students), or individually. Regardless of the setting, always carefully monitor student participation and effort.

Before giving M-COMP, familiarize yourself with the procedures for administering, timing, and scoring. This will enable you to pay maximum attention to the students during testing. Briefly introduce the tasks in grade-appropriate language. Tell the students that some of the tasks may be easy, while others may be more difficult, and that they are not expected to answer all the questions correctly. Finally, let the students know that they may not be able to complete the probe in the time allowed.

Students may use scratch paper, but the use of calculators, slide rulers, or any other assistive devices (except where dictated by a student's IEP) is prohibited. If there is material in the classroom that may help the students (such as posters of mathematic functions), please remove the materials from the students' view. Make sure all cell phones, beepers, watch alarms, and other unused electronics are turned off.

## Testing Students With Special Accommodations

Some examples of special accommodations include (a) increasing the amount of test-taking time for a particular student, (b) having a student practice the test beforehand, or (c) providing feedback during the testing process to a student about whether an answer is correct or incorrect. *These accommodations are changes in the way the test was standardized, and should not be allowed.* Like all standardized tests, using M-COMP probes with some students may be inappropriate because the demands of the test do not match the capabilities of a specific student. For example, because M-COMP requires pencil-paper test taking skills, students with severe motor problems may not be appropriate candidates for M-COMP use. Although the stratified sample includes students with disabilities, those students were administered the test in the standardized manner, with no special accommodations. For students with mild visual impairments, text enlargement may be an appropriate accommodation as it does not invalidate the standardized procedures.

Due to the standardized nature of M–COMP and the reporting requirements of benchmarking, special accommodations **cannot** be made during the three Benchmark periods (Fall, Winter, Spring); however, when you are monitoring progress frequently you have more latitude for allowing specialized accommodations, as long as you are comparing the student’s scores to his or her own scores only (i.e., individually referenced instead of norm-referenced) over a period of time. For example, during progress monitoring, it may be appropriate to increase test-time for a student with some motor impairment (e.g., from 8 to 12 minutes), providing this increase is kept standard through out the progress monitoring process. Please note, any comparison of those progress monitor scores to normative scores is inappropriate and you cannot base your interpretation on such a comparison.

### Conducting a Survey-Level Assessment or Off-Level Testing

Benchmark probes must always be administered at grade level; however, in some circumstances in progress monitoring, it may be appropriate to administer the M–COMP probes from a grade other than the student’s actual grade level in addition to the student’s grade-level probes. This process is called survey-level assessment (SLA) or off-level testing. Briefly, an individual student is tested on M–COMP probes from his or her grade-level and then at consecutively lower-grade M–COMP probes until a normative score is obtained that reflects the student’s current level of performance. The normative score is that score where the tested student performs on the probe as well as a typical student at that lower grade. For example, a Grade 5 student who performs well below average (e.g., < 10th percentile) on a Grade 5 M–COMP Benchmark probe would be tested using M–COMP Benchmark probes from successively lower grade levels (e.g., Grade 4, Grade 3) until the student’s normative score is in the average range for that lower grade level (i.e., between the 25th and 74th percentile).

This is done first and foremost as the process for writing individualized goals for frequent progress monitoring (Shinn, 2002b). Off-level testing can provide supplemental information about the severity of an achievement-performance discrepancy. A Grade 5 student whose score is within the average range of Grade 2 students on a Grade 2 M–COMP Benchmark probe has a larger performance discrepancy and a more severe mathematics achievement problem than a Grade 5 student whose performance is average when compared to an average range of Grade 4 students on a Grade 4 M–COMP Benchmark.

For more information on conducting off-level testing, see the AIMSweb manual *Progress Monitoring Strategies for Writing Individualized Goals in General Curriculum and More Frequent Formative Evaluation* (Shinn, 2002b).

## Administering the Probes

Setting the students at ease before testing begins is essential. Remember the age of the students you are testing when you present the instructions and respond to questions. The instructions are carefully worded with simple language.

When using M-COMP, keep the following in mind:

- M-COMP is not a teaching tool. You *cannot* give M-COMP probes as “practice” tests or worksheets in order to teach the material or prepare students for testing.
- M-COMP is a timed (8 minute for all grades), standardized test: Changes in the presentation of probes, use of probes (e.g., as a teaching aid), or administration time (e.g., from 8 to 10 minutes or 8 to 5 minutes) of probes violate the standardized procedure and invalidate the students’ scores on that probe.
- Encourage students to attempt each item before moving on to the next item and discourage skipping items. If you see that a student is skipping ahead without attempting each item, redirect the student to try each problem before moving to the next item.
- If a student asks a question or requests clarification, redirect him or her to any direction provided on the probe and encourage the student to do his or her best.

To administer the probes, you will need:

- the standardized administration directions found in this manual.
- A copy of the M-COMP probe.

**Note.** When you print a probe, the Answer Key is included. *Remove the Answer Key before replication and retain for your use in scoring.*

- A stopwatch or other accurate timer to monitor administration time.
- Sufficient sharpened pencils.

## Administration Directions

Read the instructions to the students verbatim. Instructions to you are in regular font. *Do not read them to the students.* The instructions you read aloud to the students are in **bold print**.

Say to the students:

**We are going to take an 8-minute math test.**

**Read the problems carefully and work each problem in the order presented, starting at the first problem on the page and working across the page from left to right. Do not skip around.**

**If you do not understand how to do a problem, mark it with an X and move on. Once you have tried all of the problems in order, you may go back to the beginning of the worksheet and try to complete the problems you marked.**

**Although you may show your work and use scratch paper if that is helpful for you in working the problems, you may not use calculators or any other aids.**

**Keep working until you have completed all of the problems or I tell you to stop.**

**Do you have any questions?**

Answer any questions the students may have, then hand the students their probes, and say:

**Here are your tests.**

**Put your name, your teacher's name, and the date on each page in the space provided, then turn over the test.**

**Do not turn the test back over or start working until I tell you to begin.**

Allow students time to write their information on the probe, then say:

**Begin.**

If a student asks a question or requests clarification, redirect him or her to the probe and say:

**Read the directions again, and work the problem as best you can.**

**If you still do not understand the problem or are unable to work it, you may move on to the next question.**

If you see that a student is skipping ahead without attempting each item, provide the following direction:

**Try to work each problem. Do not skip around.**

When the 8 minutes have elapsed, say:

**Stop and put down your pencil.**

If a student(s) continues to work, restate:

**Stop working now and put down your pencil.**

At this time, collect the probe(s) and proceed to scoring.

## Scoring Guidelines

M-COMP uses the same streamlined scoring system used with the AIMSweb Mathematics Concepts and Applications (M-CAP), released in fall 2009. Rather than scoring based on correct digits and partial credit, as in the M-CBM and M-CBM2, M-COMP scoring assigns a point value based on difficulty of 1, 2, or 3 to each item. Within each grade, the point value for a given item remains the same (i.e., if the first item is valued at 1 point on the Fall benchmark, it is valued at 1 point for every other benchmark and progress monitoring probe for that grade). This method minimizes scoring time, maximizes sensitivity to growth, controls for students who skip to the “easiest” items, and ensures the psychometric soundness of the process. Although the total points available vary modestly across grades, within a grade each probe has the exact same total point value. Table 3.1 presents the total point value per probe, per grade.

**Table 3.1 Total Point Value by Grade**

Grade	Maximum Points
1	48
2	50
3	68
4	73
5	76
6	74
7	70
8	80

## How to Score the M-COMP

Each probe file includes an Answer Key. The answers provided on the Keys are the target answers for each item on the probe, along with the point value of that answer. Scoring is a straightforward process: Circle the point value if the student's answer is correct, or circle zero if the answer is incorrect. You then simply add up the value of the correct answers to obtain the total score for the probe. Figure 3.1 presents an example of a scored Answer Key for Grade 4.

Grade 4, Probe 1 Answer Key							
Item No.	Answer	Correct	Incorrect	Item No.	Answer	Correct	Incorrect
1.	31	1	0	20.	$\frac{7}{9}$	3	0
2.	4	2	0	21.	11.9	2	0
3.	15	1	0	22.	63	1	0
4.	648	2	0	23.	7	2	0
5.	28	1	0	24.	13	2	0
6.	205	1	0	25.	$\frac{3}{7}$	3	0
7.	393	2	0	26.	342	2	0
8.	64	1	0	27.	2.1	3	0
9.	357	2	0	28.	1014	1	0
10.	18	2	0	29.	1009	2	0
11.	478	1	0	30.	18	3	0
12.	186	2	0	31.	6748	2	0
13.	310	1	0	32.	2.9	3	0
14.	12	1	0	33.	1637	3	0
15.	140	1	0	34.	677	2	0
16.	30	1	0	35.	$\frac{9}{10}$	3	0
17.	9	2	0	36.	7627	2	0
18.	14.3	3	0	37.	$\frac{2}{5}$	3	0
19.	120	1	0	38.	30 r1; 30.25; 30 $\frac{1}{4}$	3	0
Subtotal 1				Subtotal 2			
20				13			
TOTAL = Subtotal 1 + Subtotal 2				33			

**Figure 3.1 Scored Answer Key**

## M-COMP Scoring Examples

The biggest challenge in scoring is determining what to do if an answer deviates from the one provided on the Answer Key, but may still be correct. The criteria used to decide when alternate answers are or are not acceptable are based on best practices and professional judgment. The primary goal is to determine if the answer reflects an understanding of the task presented. Although the provided Answer Keys present some alternate acceptable answers, the keys are not exhaustive. If a student's answer is correct, then score as correct, regardless of whether or not the answer is in the key. Figure 3.2 shows an example of a student who has presented an answer as a decimal.

- 10** Write the fraction in lowest terms

$$\frac{5}{10} =$$

$$\frac{1}{2} = .50$$

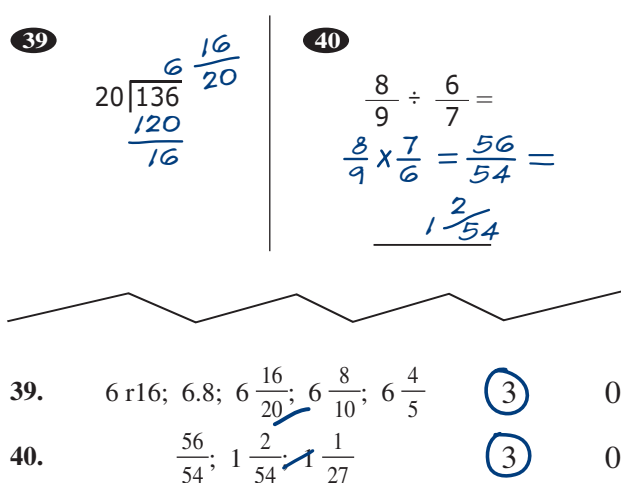
Grade 5, Probe 3 Answer Key

Item No.	Answer	Correct	Incorrect	Item No.	Answer	Correct	Incorrect
1.	128	①	0	21.	79	①	0
2.	4	①	0	22.	$\frac{63}{100}$	2	①
3.	$\frac{1}{8}$	②	0	23.	16 r3; 16.6; $16\frac{3}{5}$	3	①
4.	143	2	①	24.	5	3	①
5.	820	①	0	25.	396	①	0
6.	$\frac{1}{12}$	③	0	26.	$\frac{9}{13}$	①	0
7.	70	①	0	27.	$\frac{21}{31}$	②	0
8.	0.95	①	0	28.	45	③	0
9.	19	②	0	29.	$\frac{1}{12}$	①	0
10.	$\frac{1}{2}$	②	0	30.	$\frac{1}{15}$	③	0
11.	10 r3; 10.6; $10\frac{3}{5}$	②	0	31.	53 r3; 53.5; $53\frac{3}{6}$ ; $53\frac{1}{2}$	3	①
12.	3883	②	0	32.	33	③	0

Figure 3.2 Correct Answer Not on the Answer Key

A number of items on the M-COMP probes (Grades 4–8) result in responses that can be reduced to a simpler form. If the instructions do not specifically require that the student write the answer in the lowest terms, you may receive different correct answers. For these items, there may be a range of acceptable responses provided on the Answer Key. You will also find items, such as division items, where the correct answer can be presented with a remainder. Depending on your curriculum, it may be appropriate for the student to present this remainder as a decimal, fraction, or with an “r” followed by the remainder. Figure 3.3 shows an example of each of these item types, found in Grade 7.

**Note.** “Lowest terms” can be used interchangeably with “reduce,” “simplify,” and similar terms, depending on which term is preferred in your school’s curriculum. If your school uses the term “reduce,” tell students that when they see the instruction “lowest terms,” it means to reduce.



**Figure 3.3 Range of Acceptable Answers**

Credit may be given for a clearly correct response conveyed in a manner other than the one indicated; this is where you must rely on best practices and professional judgment.

The rest of this section presents examples of the most common variations of correct and incorrect answers seen in the national field-test sample, as well as examples of answers that require the judgment in evaluating correctness. Also included are examples of the types of issues that impact scoring decisions, including but not limited to, problems with legibility, reversed numerals, crossed-out responses, and overcorrection.



The scoring for grades 1–3 is straightforward. The problems are basic computation and number-sense questions. There is not much variability between what is correct and incorrect. At grade 4, computation with fractions and decimals are presented and it is here that some ambiguity begins to present itself.

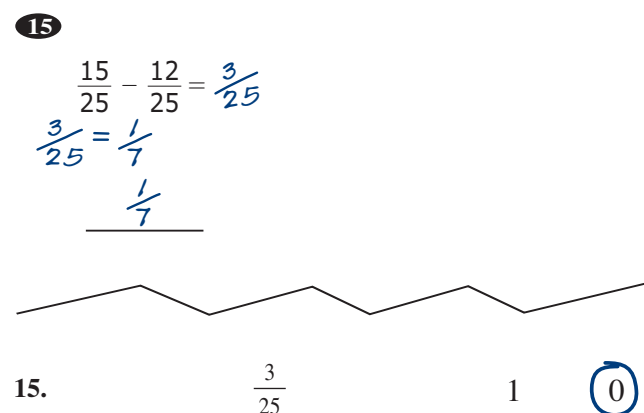
The examples in this section are not exhaustive, but representational of student responses in the national field-testing sample. Use them to guide your professional judgment when determining the correctness of answers that deviate from the correct responses identified on the Answer Key.

### Checklist for Determining Credit

- \_\_\_\_\_ Does the student's response match the answer (or alternate answers) provided on the Answer Key?
- \_\_\_\_\_ Does the student's answer represent an alternate correct answer that is not provided on the Answer Key?
- \_\_\_\_\_ Does the answer reflect an understanding of the task type?

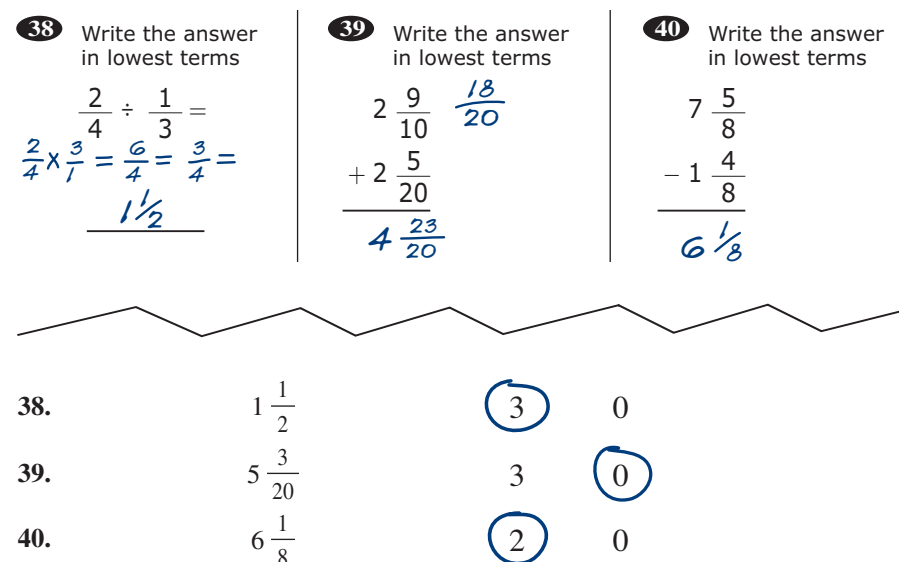
These are important questions because they reflect the basic purpose of benchmarking and progress monitoring—to determine if students are acquiring the skills required to complete the basic computational tests presented on the probes. If you encounter a scoring dilemma that is not covered in these pages, use your professional judgment in deciding the score. There is no partial scoring, so it is important to make consistent decisions when scoring potentially ambiguous answers.

For the majority of problem types, there is little deviation in acceptable answers, and where there is, it will be in the method of presentation (e.g., 0.50 versus .5). In problems where the target answer is a fraction, some students may choose to reduce the answer even when instruction to do so has not been given. Generally, as long as the reduction is performed properly, and the reduction is correct, give the student credit for the answer. This becomes tricky when a student initially provides the correct response and then makes an error in reducing. The final answer presented is what you score, so it is possible for a student to “overcorrect” into error (see Figure 3.4).



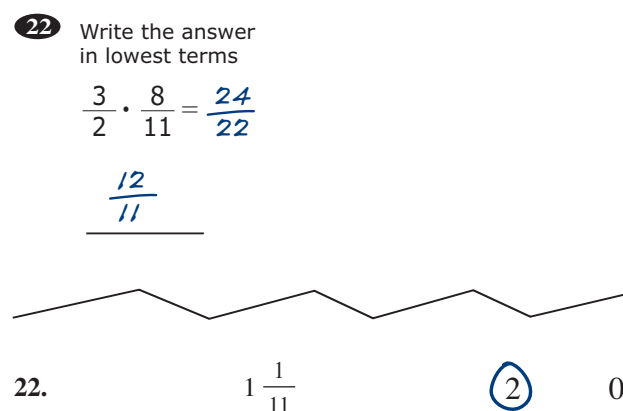
**Figure 3.4 Overcorrection Into Error**

Generally, when a specific type of target answer is required, such as an improper fraction or a mixed number, that target is requested in the directions, such as in Figure 3.5, wherein a mixed number is specifically requested as the answer.



**Figure 3.5 Targeted Answer in Directions**

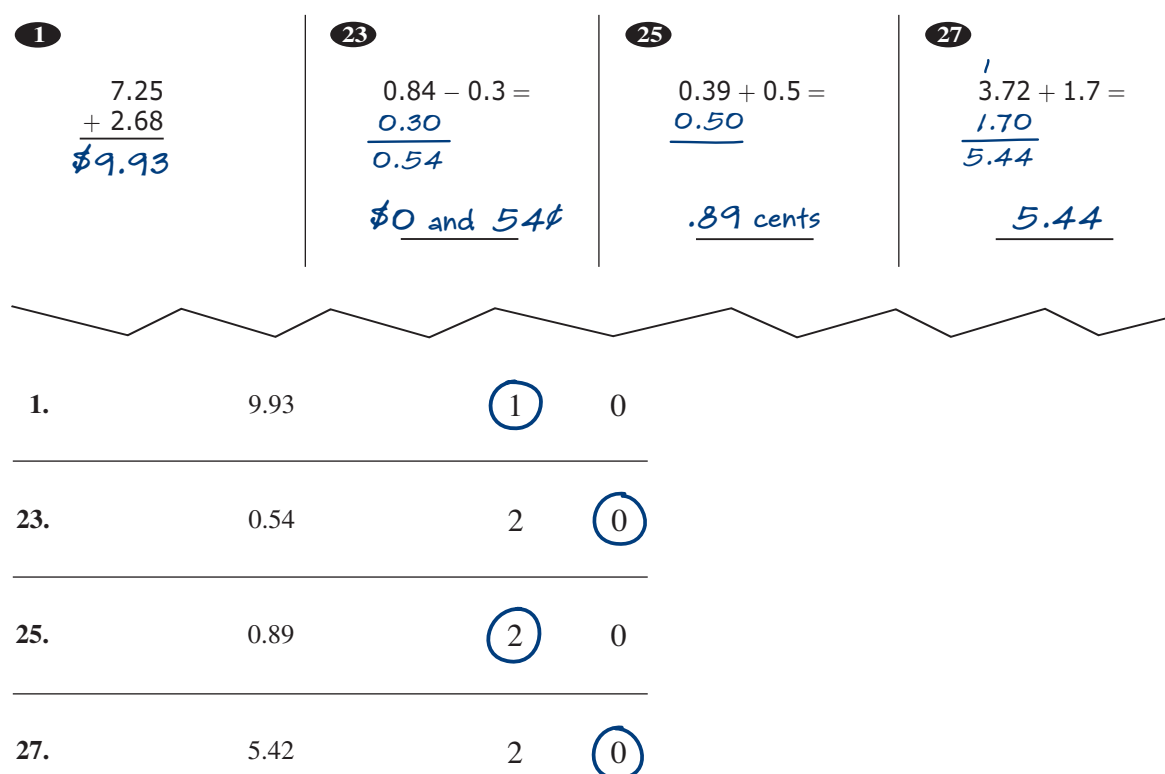
Sometimes, however, the directions may be open to interpretation, such in items where the student is instructed “Write the answer in the lowest terms.” On certain items, such as Grade 8, Item 22, the target answer is a mixed number, but the student provided an improper fraction (see Figure 3.6).



**Figure 3.6 Mixed Number for Grade 8, Item 22**

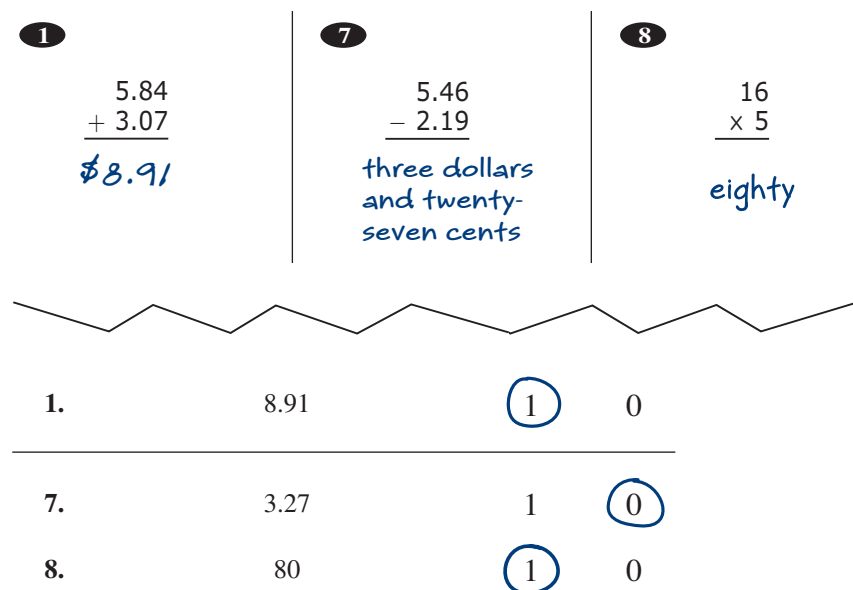
Although the target answer is a mixed number, and that is what is presented on the Answer Key, a small, but significant number of students in our national field testing provided a reduced improper fraction as an answer, as did the student in this example. After discussion with our experts, it was agreed that the nonspecific “lowest terms” could be understood by some students to be the reduced improper fraction, rather than the reduced mixed number, to be the lowest terms. For that reason, you may score a properly reduced improper fraction as correct. Because the target is that the student know how to both work the problem and reduce to a mixed number, we recommend that you provide that feedback to any student who provides an improper fraction, particularly if that student has also shown difficulty with items specifically requesting a mixed number as a response.

Another practice noted in the national field-testing sample was that, at certain grades, students added the \$ symbol to items that require addition or subtraction of decimals. Because of its prevalence, this issue was also discussed with math experts, and the decision was made that if the numerals and the decimal placement were correct, credit would be given. This issue is a bit more complicated however, when students write out the answer to the question without the decimal, relying solely on the symbols to denote the difference between the numerals preceding and following the missing decimal. Figure 3.7 presents examples of correct and incorrect answers in this area.



**Figure 3.7** Answers to Decimal Problems

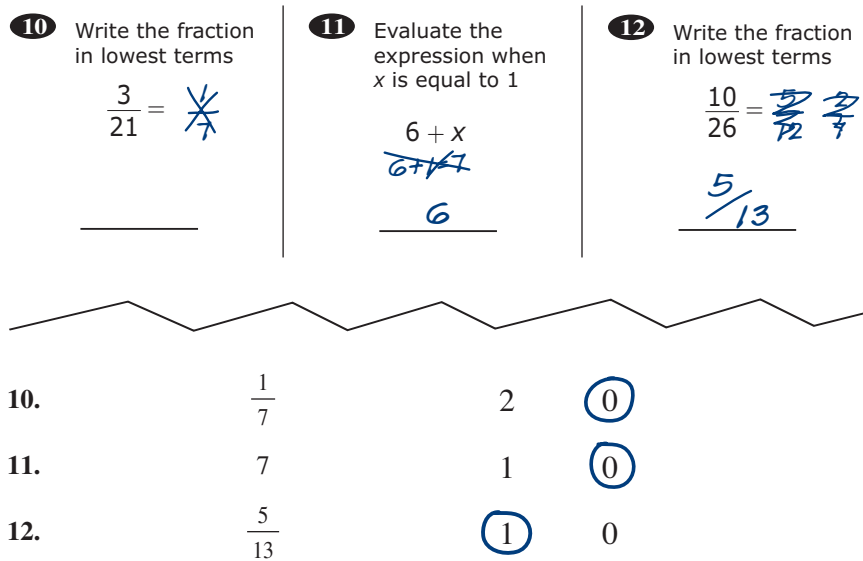
In all grades we found examples of students occasionally spelling out the answers to the problems. If the answer to a problem is the number 3 and a student writes in *three*, you may give credit. If the problem is one working with decimals and the answer is 2.5 and the student writes *two point five*, you may give credit. If, however, the student responds with *two dollars and fifty cents*, the answer is incorrect because the answer skirts the issue of decimal placement. Figure 3.8 presents an example of correct and incorrect answers from Grade 6.



**Figure 3.8 Written Answers**

Other examples of scoring issues are crossed-out answers, illegible answers, reversed numbers, or rotated numbers in answers.

- Crossed-out answers: If a student shows his or her work, but then crossed or X-ed out the problem *without* placing the answer in the blank, the item is incorrect and receives no credit. If the student has crossed out the problem, but then returned to the item and placed an answer in the blank, score the item based on whether or not the answer placed in the blank is correct. See Figure 3.9.



**Figure 3.9 Crossed-Out Answers**

- Illegible, reversed, or rotated numbers: When students write answers that have illegible, reversed, or rotated numbers, it is important to keep in mind the intent of using M-COMP probes—to determine a student’s understanding of the task and progress throughout the school year. Problems with legibility are common, particularly with the younger grades, and students identified as having specific learning challenges may have issues with reversing numbers and letters. Figures 3.10 through 3.12 provide examples of such responses.
- If the response is hard to read, but can be determined, score the answer as correct.

21	$\begin{array}{r} 6 \\ \times 7 \\ \hline 42 \end{array}$	22	$\begin{array}{r} 2 \\ 8 \overline{)16} \end{array}$	23	$\begin{array}{r} 4 \\ \times 4 \\ \hline 16 \end{array}$	24	$\begin{array}{r} 309 \\ 224 \\ + 410 \\ \hline 943 \end{array}$
21.	42	1	0				
22.	2	3	0				
23.	16	2	0				
24.	943	2	0				

**Figure 3.10 Difficult-to-Read Response**

- If the response is too illegible to determine with confidence, score it as incorrect. If the response is reversed, but the digit the student intended is obvious, score it as correct.

<b>1</b> $\begin{array}{r} 2 \\ + 3 \\ \hline 5 \end{array}$	<b>2</b> $0 + 7 = 7$ $\underline{7}$
---	--

---

1.	5	①	0
2.	7	①	0

**Figure 3.11 Reversed Numbers With Intended Number Obvious**

- If the response is rotated and you cannot easily determine what digit the student intended, score it as incorrect.

<b>25</b> $\begin{array}{r} 15 \\ + 13 \\ \hline 20 \end{array}$	<b>26</b> $\begin{array}{r} 7 \\ - 6 \\ \hline 1 \end{array}$	<b>27</b> $\begin{array}{r} 16 \\ - 10 \\ \hline 09 \end{array}$
---	--	---

---

25.	28	③	0
26.	1	②	0
27.	6	3	①

**Figure 3.12 Rotated Numbers With Intended Number Indeterminable**

A final note on scoring: Use your professional judgment in determining whether or not to give a student credit for an answer that deviates from the answer provided on the Answer Key. If the answer is mathematically correct, shows an understanding of the operation being assessed, and is consistent with the manner in which your curriculum treats that operation, then the student should get credit for the answer. When students present non-target responses, such as adding money symbols, writing out the answer as words not numbers, or providing a reduced improper fraction where a mixed number is the target, after you have scored the answer as correct, discuss the item with the student so he or she understands what is expected in the future, as continuing with certain nonstandard styles could inadvertently lead to errors on other probes.

## Reporting

---

The next step in the process is reporting your data in the AIMSweb reporting system. First, log into your school's AIMSweb account. On the opening screen there are tabs along the top and down the left side. Click on **Report** in the row of tabs along the top. At the Report page, there are tabs across the top and down the left side of the page. The row of tabs across the top represent the type of information you can report. In this case, click Mathematics. After you choose Mathematics, choose the level of the information you want to report from the tabs down the left side: Customer, District, School, Grade, or AIMSweb.

### Generating Student Reports

The most common types of reports used are the Individual Student Report, Pathway Report, and Email Report.

#### Creating an Individual Student Report

If you have entered student scores, they are listed under the column headings for each General Outcome Measure.

Select **Mathematics** from the gray tabs.

Select **M-COMP** from the corresponding radio buttons.

Click on a student's score to view the student's Individual Report.

**Note:** If you click on a column heading (e.g., RBP), an **Individual Student Report** is generated for all of the students in the classroom.



## Pathway Report

The Pathway Report for every student in the classroom is located under the **Pathway** column heading. The **Pathway Report** displays student scores for every Benchmark Period and every General Outcome Measure. Click the **Pathway** column heading to generate reports for all the students in the classroom.

**Note:** You will need *Adobe Acrobat Reader* to view and print the reports. If you do not have *Adobe Acrobat Reader* installed on your computer, download and install the latest version free of charge from <http://get.adobe.com/reader/>.

## Emailing a Report

All reports listed in this section can be generated as PDFs and emailed from the AIMSweb system.

You may email the currently selected report by clicking the **Email Report** button.

The **Email Report** screen is displayed. Enter:

Your email address (displayed by default according to the email address that was originally entered with the user information)

Recipient email address

Subject (displays a default subject)

Enter your message in the Short Message window.

Click **Send** to send the email, or click **Cancel** to return to the currently selected chart.

Click the **Back** button to return to the reports page.



# Section 4

## Developing and Standardizing the Test

### Creating the New Blueprint

The first major component and task of the M-COMP revision was determining the blueprint for each grade. The AIMSweb content team engaged internal mathematics expertise as well as multiple nationally recognized RTI and mathematics experts in the development of the new blueprints for each grade (1–8). Upon finalization of blueprints, anchor probes were developed for each of the grades. Once the anchor probes were finalized, each probe was sent to the RTI and mathematics experts, along with a team of professional educators for an additional round of input and analysis. Once all of the data were aggregated, the AIMSweb content team used the collective analyses and made final adjustments to the probes.

**Table 4.1 Domains Evaluated With M-COMP by Grade**

	1	2	3	4	5	6	7	8
Column Addition	✓	✓	✓					
Basic Facts	✓	✓	✓	✓	✓	✓		
Complex Computation	✓	✓	✓	✓	✓	✓	✓	
Fractions				✓	✓	✓	✓	✓
Decimals				✓	✓	✓	✓	✓
Reducing					✓	✓	✓	
Percentages					✓	✓	✓	✓
Conversions					✓	✓	✓	✓
Expressions						✓		
Integers						✓	✓	✓
Exponents							✓	✓
Equations							✓	✓

### Item Development

Each item was individually developed by professional test developers and mathematics experts. The development of each item was based on the blueprint by grade-level and domain-specific criteria. Each item was field tested prior to the final selection of probes. Only items with field-test data that meet specific acceptance criteria are considered for use in publication (see National Field Test section for details).

## Item Review

Following extensive internal review, revision, and pilot testing, a team of mathematics experts reviewed each item and individual probe for each of the grade levels (1–8) and provided detailed reports on their reviews. The experts were asked to evaluate myriad aspects of each item and probe to ensure the highest level of quality standards were met. The following is a list of some of the specific tasks performed and components reviewed by the experts.

- Detailed item reviews were performed to ensure no errors existed in the overall structure of the items (stem, format, etc.).
- Each item was re-worked to ensure accuracy of item function and the answer key was checked against the associated answer key.
- Item types were evaluated for equivalency and were placed consistently across probes.
- Item sequencing was evaluated to ensure clueing and other anomalies did not occur.
- Upon completion of their reviews, mathematics experts provided detailed reports to AIMSweb content, who then made the appropriate adjustments.

## Test Construction

An anchor probe was developed for each grade level (1–8). Each anchor probe was constructed by selecting individual items based on multiple criteria. All items were field-tested and then evaluated based on point-biserial correlations and item difficulty. Items that did not meet psychometric criteria were removed from the item pool. Item placement on each anchor probe was based on increasing item difficulty within a domain. To ensure the sensitivity of the instrument and to maximize the amount of data collected from at-risk learners, easier items were generally placed at the beginning of each probe and more difficult items followed.

To maintain proper randomization and representation of item type by domain, items were not placed in exact order of difficulty within each probe. Multiple item types that measure the same domain in groups of three or more also were not placed in exact order of difficulty. These anchor probes then served as the templates for developing equivalent probes at each grade.

Each equivalent probe was built to replicate the item type proportions, difficulty, and item placement on the anchor probe. For example, Item 1 of the grade 1 anchor probe is a basic addition question; therefore Item 1 for each subsequent and equivalent grade 1 probe is the same problem *type* (i.e., basic addition) with a different *item* of similar difficulty and construction. Although most items are unique (with a small percentage of repetition within a grade), the intent is to assess the same learning domain with a similar question at the same numbered position on each equivalent probe.

## Major Research Stages

There were three major research stages in creating the M–COMP benchmark and progress monitoring probes: pilot studies, national field testing, and final selection for publication.

### Pilot Studies

The primary goal of the pilot stage was to produce anchor probes with the desired content coverage and psychometric properties for which equivalent probes could be generated for national field-testing. This stage of development focused on issues such as item content and relevance, adequacy of scale floors and ceilings, appropriate time limits at administration, scoring criteria, and other relevant psychometric properties. There were three pilot studies.

**Pilot 1:** In the initial pilot, 16 students were tested to ensure that all directions for administration and individual item directions were clear and grade-level appropriate, that no issues existed with item progression, and that all items were functioning properly. Qualitative and quantitative feedback were used to make minor adjustments to prepare for Pilot 2.

**Pilot 2:** To determine alternate-form reliability and the time limits for administration at each grade level, one anchor probe and two clone probes were administered to a group of 337 students. The progress students made toward completing the probes was marked for each of the three probes at each grade at varying time points. A correlation analysis was conducted at each grade level to determine the most appropriate amount of time necessary for the test to maintain reliable discriminability. Again, qualitative and quantitative feedback was used to make adjustments to prepare for Pilot 3.

**Pilot 3:** Pilot 3 administration was untimed. The anchor probes were group administered to 444 students. The intent of this study was to extend the collection of item-specific and probe-level data from Pilots 1 and 2 to further evaluate the performance of all items. Multiple criteria, including point-biserial correlation coefficients and  $p$  values, were used to evaluate the items. Split-half correlation and Cronbach's alpha were used to evaluate the reliability of the anchor probe.

### National Field Test

A national field-test edition of the M–COMP assessment was developed. Data were obtained from a sample of 7,703 students, representing some key population demographic variables such as grade, sex, race/ethnicity, socioeconomic status, and geographic region. See Appendix A for the demographic information for the national sample.

For national field testing, 45 probes (including the anchor probe for each grade) were administered. Given the administration time limit (8 minutes), and to avoid an accelerated practice effect from repeatedly answering questions on the same domain, only a single set of six probes (including five M–COMP probes and either one off-level M–COMP probe or an M–CBM/M–CBM2 probe) was administered to each student. Twenty-two sets of probes were assembled for each grade. In each set, the anchor probe was always administered first. The remaining four M–COMP probes were administered in counter-balanced order, with half of the

sample receiving an M-CBM or M-CBM2 probe, that was placed either after the first (anchor) probe or the third M-COMP probe.

**Table 4.2 National Field Testing Item and Probe Count by Grade**

Grade	Item Count by Probe	Probe Count by Grade
Grade 1	28	45
Grade 2	28	45
Grade 3	37	45
Grade 4	38	45
Grade 5	39	45
Grade 6	40	45
Grade 7	40	45
Grade 8	40	45

### Finalizing and Selecting Probes

Multiple criteria were used to select the most psychometrically-sound equivalent probes. Pearson's product-moment correlation coefficient was examined to assess the consistency of probes within each grade. The average correlation coefficients for each probe with other probes in the set are reported in Table A.6. To evaluate the internal consistency of the probes, Cronbach's alpha and split-half reliability were examined (see Table A.6). Probe selection was based on the evaluation of these statistical properties and the comparison of the probe mean scored to the aggregated mean for each grade. Analysis of the confidence interval at the 99% level using the standard error of measurement (*SEM*) showed no statistically significant difference among the final selected probes, which were statistically equivalent to each other in the grade. The aggregated means and *SEMs* are also reported in Table A.6.



## ***Appendix A***

### ***Technical Adequacy and Data Tables***

## A.1 Demographic Characteristics of the Sample by Grade and Geographic Region

		Geographic Region							
		Northeast		Midwest		South		West	
Grade	N	%	N	%	N	%	N	%	N
1	202	22.0	229	24.9	423	46.0	65	7.1	919
2	165	16.9	272	27.9	478	49.0	61	6.3	976
3	195	20.1	172	17.7	530	54.6	74	7.6	971
4	192	21.0	97	10.6	568	62.0	59	6.4	916
5	178	17.0	151	14.4	655	62.5	64	6.1	1,048
6	96	9.8	372	37.9	440	44.9	73	7.4	981
7	87	9.2	277	29.3	436	46.2	144	15.3	944
8	92	9.7	356	37.6	350	36.9	150	15.8	948
Total	1,207	15.7	1,926	25.0	3,880	50.4	690	9.0	7,703

Note. Row percentages may not sum to 100 due to rounding.

## A.2 Demographic Characteristics of the Sample by Grade and Community Type

Grade	Community Type						Total
	Urban		Suburban		Rural		
	N	%	N	%	N	%	
1	248	27.0	485	52.8	186	20.2	919
2	203	20.8	542	55.5	231	23.7	976
3	252	26.0	502	51.7	217	22.3	971
4	235	25.7	490	53.5	191	20.9	916
5	240	22.9	644	61.5	164	15.6	1,048
6	151	15.4	724	73.8	106	10.8	981
7	63	6.7	707	74.9	174	18.4	944
8	67	7.1	802	84.6	79	8.3	948
Total	1,459	18.9	4,896	63.6	1,348	17.5	7,703

Note. Row percentages may not sum to 100 due to rounding.



### A.3 Demographic Characteristics of the Sample by Grade and Sex

Grade	Sex				Total
	Female		Male		
	N	%	N	%	
1	489	53.2	430	46.8	919
2	545	55.8	431	44.2	976
3	544	56.0	427	44.0	971
4	513	56.0	403	44.0	916
5	562	53.6	486	46.4	1,048
6	504	51.4	477	48.6	981
7	504	53.4	440	46.6	944
8	537	56.6	411	43.4	948
Total	4,198	54.5	3,505	45.5	7,703

Note. Row percentages may not sum to 100 due to rounding.

### A.4 Demographic Characteristics of the Sample by Grade and Race/Ethnicity

Grade	Race/Ethnicity												Total
	African American		American Indian		Asian		Hispanic		White		Other <sup>a</sup>		
	N	%	N	%	N	%	N	%	N	%	N	%	
1	96	10.4	19	2.1	19	2.1	237	25.8	536	58.3	12	1.3	919
2	86	8.8	18	1.8	15	1.5	247	25.3	595	61.0	15	1.5	976
3	87	9.0	12	1.2	15	1.5	269	27.7	579	59.6	9	0.9	971
4	88	9.6	13	1.4	7	0.8	260	28.4	544	59.4	4	0.4	916
5	99	9.4	35	3.3	5	0.5	225	21.5	679	64.8	5	0.5	1,048
6	89	9.1	25	2.5	11	1.1	152	15.5	693	70.6	11	1.1	981
7	37	3.9	28	3.0	10	1.1	234	24.8	626	66.3	9	1.0	944
8	58	6.1	28	3.0	23	2.4	187	19.7	646	68.1	6	0.6	948
Total	640	8.3	178	2.3	105	1.4	1,811	23.5	4,898	63.6	71	0.9	7,703

Note. Row percentages may not sum to 100 due to rounding.

<sup>a</sup> Includes Alaska Natives, Pacific Islanders, Native Hawaiians, and all other groups not classified as African American, American Indian, Asian, Hispanic, or White.

## A.5 Demographic Characteristics of the Sample by Grade and Median Family Income

Grade	Median Family Income Level						Total N
	Low		Middle		High		
	N	%	N	%	N	%	
1	462	50.3	150	16.3	307	33.4	919
2	517	53.0	115	11.8	344	35.2	976
3	381	39.2	267	27.5	323	33.3	971
4	327	35.7	364	39.7	225	24.6	916
5	366	34.9	433	41.3	249	23.8	1,048
6	234	23.9	328	33.4	419	42.7	981
7	293	31.0	371	39.3	280	29.7	944
8	218	23.0	356	37.6	374	39.5	948
Total	2,798	36.3	2,384	30.9	2,521	32.7	7,703

Note. Row percentages may not sum to 100 due to rounding.

## A.6 Descriptive and Reliability Statistics by Grade

Grade	Mean <sup>a</sup>	SD <sup>b</sup>	SEM <sup>c</sup>	r <sup>d</sup>	Split-Half <sup>d</sup>	Alpha <sup>d</sup>
1	36.0	12.8	4.02	.86	.89	.87
2	37.9	11.4	4.04	.82	.85	.82
3	51.2	17.6	4.67	.89	.90	.89
4	51.3	17.1	5.56	.85	.91	.87
5	33.7	20.3	5.67	.89	.93	.91
6	32.5	17.9	4.79	.89	.89	.89
7	34.8	19.4	5.06	.90	.92	.91
8	30.2	18.9	5.59	.88	.92	.90

<sup>a</sup> Weighted average.

<sup>b</sup> Pooled standard deviation.

<sup>c</sup> The SEM for each probe was calculated based on the average correlation coefficient and the actual standard deviation of the raw score for the probe. The average SEM for the grade was calculated by averaging the squared SEMs for each probe and obtaining the square root of the result.

<sup>d</sup> The average reliability coefficients were calculated using Fisher's z transformation.

### A.7 Correlations of AIMSweb M-COMP Scores With Group Mathematics Assessment and Diagnostic Evaluation (G-MADE) Scores by Grade

Grade	M-COMP		G-MADE		$r^a$	Total <i>N</i>
	Mean	<i>SD</i>	Mean	<i>SD</i>		
1	36.0	12.2	17.9	5.6	.84	98
3	45.5	14.2	15.9	5.7	.73	98
8	28.4	13.6	13.4	3.8	.76	54

<sup>a</sup>The average correlation coefficients across both administration orders were calculated using Fisher's *z* transformation.



**Aim Line**—Graphical representation of student's expected rate of progress.

**Benchmark Probe**—Administered to all students in Fall for screening, Winter and Spring for progress monitoring and program evaluation.

**Box and Whisker Plot**—A graphical representation of a set of scores divided into four parts using quartiles and median.

**Content Validity**—A type of validity that relates to how adequately the content of a test represents a specified body of knowledge, and to how adequately subjects' responses represent knowledge of the content.

**Correlation**—A measure of the strength and direction of the relationship between two sets of variables. (See Correlation Coefficient).

**Correlation Coefficient ( $r$ )**—A statistic ranging between 0 and 1 that indicates the degree and direction of relationship between two variables. The strength of the relationship is indicated by the values of the coefficients (with greater values indicating stronger relationships). The direction of the relationship is indicated by either a positive sign (+) representing a positive relationship in which variables tend to increase or decrease together, or a negative sign (–) representing an inverse relationship between variables.

**Individually Referenced**—Comparing a student's scores to his or her own scores (instead of norm-referenced) over time.

**Internal Consistency**—A type of test score reliability indicating the degree of correlation among item responses within each separate part of a test.

**Internal Structure Validity**—A type of validity involving the degree to which relationships among test items and test components conform to what the test is intended to measure.

**Line of Best Fit**—A line drawn through individual data points.

**Longitudinal Tracking**—The tracking of particular data (e.g., mean entering scores) over a long period of time to establish trends.

**Mean ( $M$ )**—The average of a set of scores computed by adding all of the scores together and then dividing by the total number of scores.

**Median**—The middle value in a distribution of score with 50% of the scores lying below it (i.e., the 50th percentile).

**Meta-Analysis**—A method of research that analyzes the results of several independent studies by combining them to determine an overall effect or the degree of relationship between variables.

**N count ( $N$ )**—The total number of individuals who make up a sample (e.g., the number of students that took a probe).

**Normative Score**—Score on a scale where the tested student performed as well as the typical student at that grade.

**Normative Sample/Norm Group**—The group of individuals (sample) earning scores on a test whose score data are used to determine scaled scores and/or percentile ranks.

**Off-Level Testing (Survey-Level Assessment)**—Administering probes from a grade other than the student's actual grade level in some circumstances, in addition to the student's grade-level probes until a normative *score* is obtained that reflects the student's current level of performance.

**Percentile Rank (PR)**—A whole number between 1 and 99 that represents the proportion of individuals from the normative sample who earned lower than a given score on a test.

**Progress Monitoring Probe**—Administered to students identified as at risk, to monitor student progress and program evaluation. Frequency to be determined by teacher or IEP, if applicable.

**Predictive Validity**—A type of validity based on how accurately test data (e.g., admission test scores) are able to predict criterion measures obtained at some later time (e.g., a grade point average earned after admission).

**Predictor Variable**—A variable that occurs prior to any intervention (e.g., scores on an admission test) that is used to predict some subsequent outcome (e.g., a grade point average earned after admission). (See Variable.)

**Rate of Improvement (ROI)**—The rate of improvement measures the progress the student makes per progress monitoring event since the start of an intervention program. To determine the ROI, divide the amount of progress per event by the number of weeks of intervention.

**Raw Score (RS)**—The number of items answered correctly by a candidate on a test.

**Reliability**—An estimate of the dependability of test scores in terms of the degree of consistency between measures of the test (e.g., comparisons of administrations of a test over time, or comparisons of items within a test).

**Reliability Coefficient**—A correlation statistic (usually ranging from 0 to 1) that measures the degree to which test scores are free of measurement error. (See Standard Error of Measurement.)

**Standard Deviation (SD)**—A measure of the variability of test scores in terms of how spread out scores are from one another in a normative sample distribution.

**Standard Error of Measurement (SEM)**—An estimate (based on group data) of the variation in scores earned on repeated administrations of the same test by the same individual.

**Trend Line**—A graphical representation of an individual student's rate of improvement (ROI).

**Validity**—The extent to which a test measures what it is intended to measure. Validity refers to the extent to which test scores (or other measures) support appropriate interpretations and inferences regarding characteristics of a person measured (e.g., knowledge or ability) or performances other than those measured (e.g., subsequent performance or achievement).



## References

- Shinn, M. R. (2002a). *Organizing and implementing a benchmark assessment program*. Eden Prairie, MN: Edformation.
- Shinn, M. R. (2002b). *Progress monitoring strategies for writing individualized goals in general curriculum and more frequent formative evaluation*. Eden Prairie, MN: Edformation.
- Shinn, M. R. (2008). Best practices in curriculum-based measurement and its use in problems-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., pp. 243–262). Bethesda, MD: National Association of School Psychologists.

