Making Connections

February 2017

Measuring student progress and teachers' assessment of student knowledge in a competency-based education system

R. Marc Brodersen
Marzano Research

Bruce Randel
Century Analytics, Inc.

Key findings

This report examines the academic progress of elementary and middle school students enrolled in competency-based education in one Colorado district. In competency-based education, students are promoted to the next performance level once their teacher determines that they have demonstrated mastery of all the learning targets for a course. The study found that:

- A majority of students completed their math and reading performance levels in approximately one academic year, but 43–47 percent of the students who were behind their traditional grade levels completed their performance levels in three or fewer quarters, less time than it would take in a traditional education system.
- Teachers' assessments of student competencies were only weakly related to student math and reading achievement on the Transitional Colorado Assessment Program, the state's standardized test.

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

REL CENTRAL
Regional Educational Laboratory
At Marzano Research

U.S. Department of Education
Betsy DeVos, *Secretary*

Institute of Education Sciences
Thomas W. Brock, *Commissioner for Education Research*
Delegated the Duties of Director

National Center for Education Evaluation and Regional Assistance
Audrey Pendleton, *Acting Commissioner*
Elizabeth Eisner, *Acting Associate Commissioner*
Amy Johnson, *Action Editor*
Sandra Garcia, *Project Officer*

REL 2017–238

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased large-scale evaluations of education programs and practices supported by federal funds, provides research-based technical assistance to educators and policymakers, and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

February 2017

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-12-C-0007 by Regional Educational Laboratory Central administered by Marzano Research. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Brodersen, R. M., & Randel, B. (2017). *Measuring student progress and teachers' assessment of student knowledge in a competency-based education system* (REL 2017–238). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. Retrieved from http://ies.ed.gov/ncee/edlabs.

This report is available on the Regional Educational Laboratory website at http://ies.ed.gov/ncee/edlabs.

Summary

Competency-based education has received growing attention in recent years as a way to address preK–12 learning goals. In competency-based education, students are promoted to the next course of study or grade level in each subject area after demonstrating mastery of identified learning targets aligned to standards (Wolfe, 2012). By contrast, in traditional education, students earn credit for a course after spending a required amount of time (for example, one semester) in that course and meeting minimum course criteria.

Westminster Public Schools in Colorado began the transition to a competency-based education system in 2009 (Sturgis & Patrick, 2010). In the Westminster Public Schools system, courses of study are organized according to performance level rather than according to traditional grade levels. Performance levels are analogous to grade levels in a traditional education system, although students are placed in a performance level on the basis of their current knowledge and skills rather than on the basis of their age or number of years in school. Performance levels are organized around a series of learning targets that are aligned to state standards. Teachers make ongoing judgments about student proficiency by assigning scores on each learning target, and students advance to the next performance level by demonstrating proficiency in all learning targets for a particular subject area. While most students are expected to complete one performance level per year, students may progress through their performance levels at their own pace. When appropriate, students may take more or less than one year to complete a performance level.

Westminster Public Schools, a member of the Regional Educational Laboratory Central College and Career Readiness Research Alliance, asked for assistance in examining how long it takes students in the district to progress through their performance levels, especially students who are below their traditional grade level. This information would help the district and other educators interested in competency-based education to determine whether allowing students to progress through performance levels at their own pace provides students who are below their traditional grade level the opportunity to catch up and meet grade-level proficiency expectations. Westminster Public Schools also asked for assistance in examining how well teachers' ratings of student competency (learning target scores) align to external assessments of student academic achievement. Educators may use the approach described in this report to assess the degree to which teachers' judgments of student competency relate to student academic achievement measures in their own school or school district.

Using data from the Westminster Public Schools learning management system, this study examined how long elementary and middle school students took to complete math and literacy performance levels 3–8 during the 2013/14 school year. To examine the relationship between students' learning target scores and Colorado's standardized achievement test scores, a student's learning target scores within a performance level were combined to create an overall performance-level competency score for each student. The performance-level competency scores represent an aggregate measure of student competency within their given performance levels based on teachers' judgments. These performance-level competency scores were then used to predict students' scores and proficiency levels on the Transitional Colorado Assessment Program.

A majority of students completed their courses of study in math and literacy in approximately one academic year. Although a majority of students who were in a math or literacy performance level below their traditional grade level also completed their course of study in one academic year, a larger percentage of them (43–47 percent) completed their level in three or fewer quarters compared with students in a performance level at their traditional grade level (17–22 percent). These results suggest that competency-based education in Westminster Public Schools provides students who are behind academically an opportunity to complete performance levels in less time than in a traditional education system.

Students' performance-level competency scores had statistically significant and positive relationships with Transitional Colorado Assessment Program scores, but the relationships were weak. The performance-level competency scores accounted for only a small proportion (3–4 percent) of the variance in students' scores on the state achievement test. Math performance-level competency scores accurately predicted math proficiency levels on the state achievement test for 40 percent of students, and literacy performance-level competency scores accurately predicted reading proficiency levels on the state achievement test for 59 percent of students. The performance-level competency scores of students who were in a performance level below their traditional grade level were more likely to predict that their state achievement test proficiency level would be higher than it actually was. In contrast, for students above grade level, performance-level competency scores were more likely to predict that their state achievement test proficiency level would be lower than their actual level. The relatively weak relationships between performance-level competency and state achievement test scores suggest that teachers' judgments of student competency under competency-based education in Westminster Public Schools are not good predictors of academic performance, as measured by the Transitional Colorado Assessment Program.

Contents

Tables

Why this study?

As states establish rigorous academic standards for college and career readiness, new education approaches may be needed to help students learn and graduate with the knowledge and skills associated with these standards. One model that has gained attention is competency-based education, which requires students to demonstrate mastery of identified learning targets aligned to standards in order to be promoted to the next course of study or grade level in each subject area (Wolfe, 2012; see box 1 for definitions of key terms). By contrast, in traditional education, students earn credit for a course after spending a required amount of time (for example, one semester) in the course and meeting minimum course criteria. This report presents the results of a study examining how long it takes students to complete their courses of study in a school district that has adopted competency-based education and how well teachers' judgments of student competency correspond to an external assessment of student knowledge.

This report presents the results of a study examining how long it takes students to complete their courses of study in a school district that has adopted competency-based education and how well teachers' judgments of student competency correspond to an external assessment of student knowledge

A competency-based education helps ensure that students will attain competency in course content by allowing them to master all course learning objectives at their own pace. Additionally, "competency-based education explicitly recognizes that students are different —with different sets of skills, different levels of maturity and identity, and different aptitudes, interests, and family supports.… For all students to be college and career ready, one of the fundamental changes in competency education is the way students progress through the education system" (Patrick & Sturgis, 2013, p. 22). Such a system can benefit students for whom the regular classroom pace is too slow by allowing them to progress more rapidly through course content.

Box 1. Key terms

At, below, or above grade level. In Westminster Public Schools a student who is at grade level is in a performance level that corresponds to the student's traditional grade level (for example, a grade 5 student is in math performance level 5). A student who is below grade level is in a performance level below the student's traditional grade level (for example, a grade 5 student performing below the expected level in math could be receiving instruction and support in math at performance level 3 and be targeted for catch-up). A student who is above grade level is in a performance level above the student's traditional grade level (for example, a grade 5 student in math performance level 7).

Competency-based education. Competency-based education (alternatively referred to as proficiency-based, mastery-based, and performance-based education) refers to an education system in which students are promoted to the next course of study after they demonstrate competency in course content. Competency is determined through ongoing assessment by the classroom teacher, and students are allowed flexible time as necessary to master the course content.

Course of study. A grouping of students by performance level and content area, similar to a traditional classroom (for example, math performance level 5).

Cutscore. A selected point on the score scale of a test used to determine whether a particular test score is adequate for some purpose. For example, cutscores developed by the Colorado Department of Education are used to divide students' scale scores on the Transitional Colorado Assessment Program into four performance levels: unsatisfactory, partially proficient, proficient, and advanced.

(continued)

Box 1. Key terms *(continued)*

Learning management system. A computer software application that allows teachers to enter students' learning target scores and track student progress in achieving competency in each learning target.

Learning target. Defined learning targets for each performance level in each content area (for example, math performance level 5) that represent the essential skills and knowledge addressed in that course of study and are aligned to state academic standards. The number of learning targets in each performance level examined by this study (math and literacy performance levels 3–8) range from 33 to 49, with an average of 39 learning targets per performance level. An example learning target for literacy performance level 4 is "Determine a theme of a story, drama, or poem from details in the text; summarize the text."

Learning target score. Scores determined by teachers that indicate students' current level of competency for each learning target within a performance level. Scores are on a 0–4 scale in half-point increments. A score of 3.0 indicates that a student has mastered the learning target. Teachers update each learning target score for a given student as they obtain additional evidence on the student's current level of competency. Teachers can rely on a variety of evidence (such as classroom assignments, group discussions, or benchmark assessments) to determine a student's learning target score and update these scores at their discretion. (For additional detail on learning target scores, see appendix A.)

Performance level. Westminster Public Schools courses of study are organized according to performance level rather than grade. Elementary and middle school students are assigned to one of eight performance levels (levels 1–8) in each core subject area (for example, math and literacy). Performance levels are analogous to grade levels in a traditional education system, although students are placed into a performance level based on their current level of understanding of the content, rather than their age or number of years in school. A student's performance level may differ across content areas. Students complete a performance level once they receive scores of 3.0 or higher on all learning targets for that level and then advance to the next level. While most students are expected to complete one performance level per year, students may progress through their performance levels at their own pace. When appropriate, students may take more or less time than one year to complete a performance level.

Performance-level competency score. An overall measure of each student's level of competency within a particular performance level that was developed for this study by aggregating all the learning target scores students had at the time of the 2014 Transitional Colorado Assessment Program administration.

Traditional grade level. A student's chronological grade in most traditional, non-competency-based education schools, determined by age or number of years in school.

Transitional Colorado Assessment Program. Colorado's standardized achievement test. Students are assessed annually in grades 3–8 in reading, writing, and math. The test is administered to students at their traditional grade level and is aligned with Colorado Academic Standards. The test is vertically scaled and consists of common items that are administered across grade levels so that scores are comparable across grade levels. This study used only math and reading scores. Within each content area and grade level, students receive a scale score. The scale score can be used to divide student performance into four performance levels: unsatisfactory, partially proficient, proficient, and advanced.

Competency-based education may also help struggling students in two main ways. First, the ongoing student assessment that is associated with competency-based education can identify students who are struggling and who may need additional resources and support. Second, competency-based education can enable students who are in a class below their traditional grade level to catch up and meet grade-level proficiency expectations sooner than in a traditional system. This is possible because students may complete a class in less time than scheduled for the course (for example, two academic quarters instead of the four quarters scheduled for the course) provided they demonstrate mastery of all of that class's learning objectives.

The popularity of competency-based education is reflected in state and district policy initiatives. For example, almost a third of states that applied for the second round of federal Race to the Top education funds included competency-based strategies in their application (Sturgis & Patrick, 2010). Additionally, all but nine states have policies that provide schools and districts with some flexibility in awarding competency-based credits (White, 2014).

Competency-based education depends on teachers' ability to assess student knowledge

Effective competency-based education depends, in a large part, on teachers' ability to assess students' competency accurately using various types of evidence. The ability of teachers to assess student competency accurately has been identified as a critical teacher skill (Sudkamp, Kaiser, & Moller, 2012). Accurate assessment allows teachers to provide targeted feedback to a student, to inform subsequent instruction, and to determine when a student is ready to be promoted to the next course (Marzano, 2006). Marzano and Kendall (2008) have argued that teachers' assessments are expected to be particularly accurate in a competency-based system, where students do not progress within a class or to the next level until they have demonstrated proficiency based on assessments that are aligned to standards that are broken down into clearly defined proficiencies outlining an expected learning progression. While research has examined the relationship between teacher judgment and other measures of student achievement (Sudkamp et al., 2012), no such studies have been conducted in competency-based systems.

Effective competency-based education depends, in a large part, on teachers' ability to assess students' competency accurately using various types of evidence

Westminster Public Schools' competency-based system

Westminster Public Schools in Colorado began transitioning to competency-based education in 2009 (Sturgis & Patrick, 2010). Courses of study are organized according to performance levels rather than traditional grade levels. Each performance level is structured around a set of district-defined learning targets that are aligned with Colorado Academic Standards and represent the essential skills and knowledge addressed in the course of study. Students are placed in the performance level that is most appropriate to their skills and knowledge. Students are promoted to the next performance level once they have attained proficiency in all of the current level's learning targets. Thus, a given classroom (for example, math performance level 5) may comprise students from multiple traditional grade levels, though generally from no more than three levels. Additionally, because students can progress through a performance level's content at their own pace, some students may progress through multiple performance levels (and thus, multiple classrooms) within a given content area during a single school year. Conversely, some students may take longer than a school year to progress through one level.

Westminster Public Schools teachers periodically assess and rate student progress in mastering the learning targets of their relevant content area and performance level on a 0.0–4.0 scale, with increments of 0.5 (adapted from the Marzano Taxonomy of Educational Objectives [Marzano & Kendall, 2008]; see appendix A). As described by the district, "To achieve mastery on any Learning Target, a learner must demonstrate what they know by engaging in a series of evidence-based formative checks (assignments, informal assessments, classwork, evidence of learning, etc.) that assess their level of understanding for each learning target. A score of 1.0 or 2.0 from these formative checks during the early learning stages designates that a student is simply at the beginning stage of learning about a new topic, not that they are failing. Whether the assessment, assignment, or activity is formative or summative, the teacher will record a score from 0 through 4 as part of the scoring record for each student. These scores become the triggers for action that determine the next appropriate instructional steps" (Adams County School District 50, 2011, p. 3).

Teachers determine learning target scores for each student based on reviews of written assignments, informal assessments, district benchmark assessments, and classroom discussions. The district does not require teachers to use a specific type or amount of evidence in assigning learning target scores. Learning target scores are entered into the district's learning management system to track student progress in mastering each learning target. Teachers update the learning target scores as they obtain additional evidence of student learning. When students receive a score of 3.0 or above on a learning target, they are considered to have mastered that target. Students are promoted to the next performance level when they have earned a 3.0 score or above on all the given performance level's learning targets. In most cases, this promotion involves moving into a new classroom.

Focus of this study

As a member of Regional Educational Laboratory (REL) Central's College and Career Readiness Research Alliance, Westminster Public Schools asked for assistance in examining how long it takes students who are in a performance level below, at, or above their traditional grade level to complete their performance levels. No previous analyses had been conducted in the district to determine how long students take to complete each performance level, in part because of recent changes in state academic standards and subsequent changes in district learning targets. Understanding the typical amount of time students take to complete their performance levels may help teachers track student progress and assist with identifying students who are taking longer than is typical and who may need additional support. This information may also help the district and other educators interested in competency-based education determine whether allowing students to progress through performance levels at their own pace provides those who are below their traditional grade level with the opportunity to catch up and meet grade-level proficiency expectations.

Understanding the typical amount of time students take to complete their performance levels may help teachers track student progress and assist with identifying students who are taking longer than is typical and who may need additional support

Westminster Public Schools also asked for an examination of how well teachers' ratings of student competency (students' learning target scores) correspond to an external measure of student academic achievement, the Transitional Colorado Assessment Program, Colorado's standardized achievement test. Teachers use a variety of evidence to judge student competency, and how well these judgments relate to other measures of student academic achievement is unknown. The results of this study may inform future professional development and district policy. For example, the district may determine based on the findings that teachers need additional professional development in classroom assessment practices.

With little empirical research available on competency-based education, stakeholders such as members of the REL Central College and Career Readiness Research Alliance, competency-based education professional development providers, researchers, and educators are eager to learn about how competency-based education is being implemented, including challenges and lessons learned. For school and district leaders implementing or contemplating implementing competency-based education, this report provides information about how a competency-based education system in one district operates. Educators can use the approach described in this study to assess how well teachers' judgments of student competency relate to student academic achievement measures in their own schools or districts. This report may also prompt school and district administrators to examine the degree to which their own systems are being used as intended.

This report provides information about how a competency-based education system in one district operates. Educators can use the approach described in this study to assess how well teachers' judgments of student competency relate to student academic achievement measures in their own schools or districts

What the study examined

This study addressed the following research questions:

- How long do students who are below, at, or above grade level take to complete Westminster Public Schools math and literacy performance levels 3–8?
- To what degree do teachers' ratings of student competency correspond to math and reading performance on the Transitional Colorado Assessment Program, the state achievement test?

Descriptive and correlational analyses were used to examine administrative data collected between 2010 and 2014 (see box 2 and appendix B). Data on students in math and literacy performance levels 3–8 were used to estimate the number of instructional days and academic quarters each student took to complete performance levels.

To determine how well teachers' ratings of student competency correspond to math and reading performance on the state achievement test, the study team calculated math and literacy performance-level competency scores for each student by aggregating the multiple learning target scores provided by teachers (see box 2). The study team examined the correlation between students' performance-level competency scores calculated at the time the state achievement test was administered and students' spring 2014 math and reading scores on the state achievement test, as well as the percentage of students whose math and reading proficiency levels were accurately predicted using their performance-level competency scores.

What the study found

The descriptive findings for the first research question detail the amount of time students took to complete a performance level in math and literacy in Westminster Public Schools. The correlational findings for the second research question describe the relationships between students' performance-level competency scores and their performance on the state achievement tests (for more detailed findings, see appendix C).

Box 2. Data and methods

Data

The Westminster Public Schools performance management system contains data for 2010 to the present. Data gathered from the system for this study included students' performance level start and completion dates, learning target scores, and 2014 Transitional Colorado Assessment Program scores. The analysis sample for the first research question on how long it took students to complete math and literacy performance levels included all Westminster Public Schools students who had completed at least one math (N = 2,104) or literacy (N = 2,086) performance level from 3 to 8 during the 2013/14 school year. The analysis sample for the second research question on how well teachers' ratings of student competency corresponded to performance on the state achievement tests included all students who took the 2014 Transitional Colorado Assessment Program test and who were in math (N = 2,388) or literacy (N = 1,702) performance levels 3–8 during the 2013/14 school year and who had enough (at least five at the time of test) learning target scores in their performance level to calculate their performance-level competency score. Both samples included students who had started a performance level prior to the 2013/14 school year and had not been in that performance level for more than two years (students who had been in a performance level for more than two years may have left the district). Sample sizes varied across the two research questions because of differences in how the samples were defined. For example, the sample for the second research question omitted students with fewer than five learning target scores, which could occur because students had recently begun the performance level or the teacher had not yet assessed student knowledge of these targets. Learning target scores were not a defining factor for the first research question sample. Student composition across all of the samples was similar: 41 percent of math students in both samples were in a performance level at grade level, while 54 percent of literacy students were at grade level for the first research question sample and 51 percent for the second. Across all samples, fewer than 10 percent of students were in a performance level above grade level.

Methods

The average amount of time students took to complete performance levels was calculated in terms of the number of instructional days and the estimated number of academic quarters. For the Westminster Public Schools 2013/14 school calendar, which contained four quarters and 173 instructional days, each quarter was estimated based on an average of 42 instructional days per quarter. Because students can start and complete a performance level at different times, these estimates may not represent the actual 2013/14 academic quarter in which a performance level was completed, but they provide an estimate of the number of quarters students took to complete a performance level. The average numbers of instructional days and academic quarters were computed separately for students who were in a performance level below, at, or above their traditional grade level. Students were not considered to have been in a performance level if they had 10 or fewer days in that level because teachers had 11 instructional days to determine whether students' placement was appropriate or whether they should be assigned to a different performance level. If students completed more than one performance level during the study period and had more than 10 instructional days in each, they were included in the calculations for each of their relevant performance levels.

To examine how teacher ratings of student competency corresponded to an external measure of student academic achievement, each student's learning target scores were

(continued)

Box 2. Data and methods *(continued)*

combined into an overall performance-level competency score for that student within each math and literacy performance level. Performance-level competency scores were calculated from all the learning target scores students had received by the time they took the state achievement test. Learning target scores from previous years were included for students who started the performance level prior to the 2013/14 school year. Literacy performance-level competency scores were calculated using only the performance levels' learning targets within the district's reading content strand, which are aligned to the state's reading assessment. All math learning target data were used to calculate math performance-level competency scores. Hierarchical linear modeling was used to estimate the correlation between students' math and literacy performance-level competency scores and state achievement test scores in math and reading. Hierarchical linear modeling takes into account that students' scores can be influenced by the school in which they are enrolled, allowing for a more precise estimate of the relationship between the two measures than a simple correlation. Separate analyses were conducted for each content area.

The Colorado Department of Education uses state achievement test scores to assign students to one of four proficiency levels: unsatisfactory, partially proficient, proficient, or advanced. A predicted proficiency level on the state achievement test was calculated for each student based on the student's performance-level competency score (see appendix B for details). Because students may be in a performance level that does not correspond to their traditional grade level, students' actual proficiency levels were based on the state assessment's cutscores that aligned to their performance level (for example, the grade 5 math proficiency level cutscores were used to determine the actual proficiency levels of all students in math performance level 5, regardless of their traditional grade). Students' predicted and actual proficiency levels were compared to determine the degree of correspondence between the two.

Additional technical details on the data and methods used in this study are provided in appendix B.

On average, students who were in courses below their traditional grade level took less time to complete their courses than students who were in courses that corresponded to their grade level

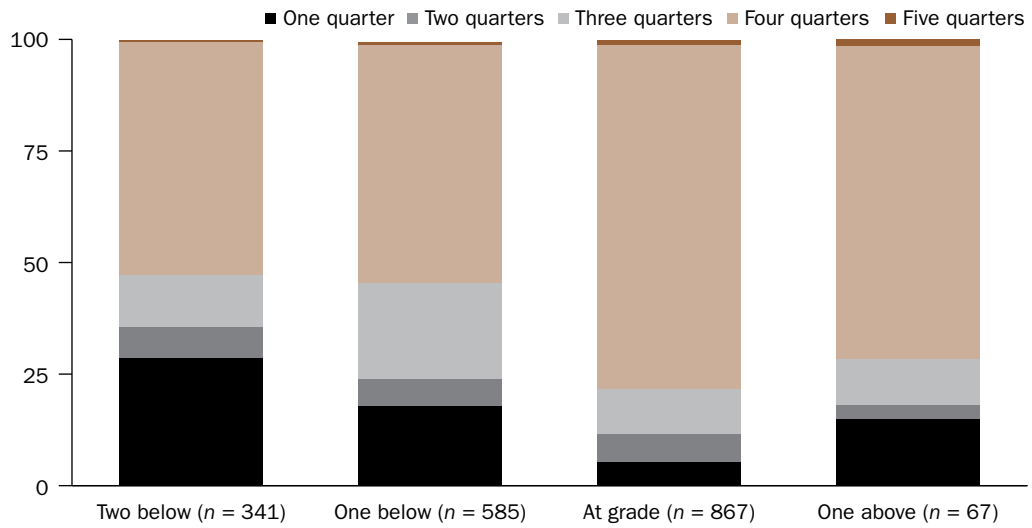
For ease of presentation, results are reported here only for students who were in a performance level between two levels below and one level above their traditional grade levels (around 88 percent of math students and 94 percent of literacy students) and who had completed a performance level in one to five quarters (approximately 97 percent of the sample). Data on all students can be found in appendix C. Among students who were in a math or literacy performance level that matched their traditional grade level, three-quarters (77 percent in math and 75 percent in literacy) completed that level in four quarters (figures 1 and 2).

Among students who were in performance levels below their traditional grade level, nearly half (43 percent for math and 47 percent for literacy) completed their levels in three or fewer quarters, faster than it would traditionally take them to advance one grade

Among students who were in performance levels below their traditional grade level, nearly half (43 percent for math and 47 percent for literacy) completed their levels in three or fewer quarters, faster than it would traditionally take them to advance one grade. A larger percentage of students whose performance levels were below their traditional grade level than of students who were at grade level completed their levels in three or fewer quarters.

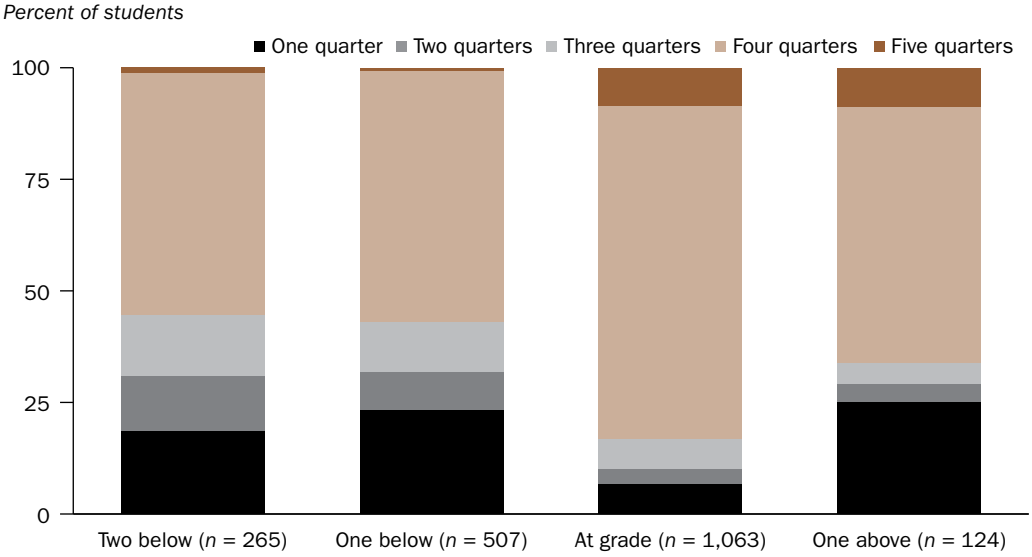
The majority of students who were in performance levels above their traditional grade level completed their levels in four quarters (70 percent for math and 57 percent for literacy).

Figure 1. Students who were below and those who were above their traditional grade level in math took fewer quarters to complete their performance levels than did students at grade level, Westminster Public Schools, 2013/14



Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

Figure 2. Students who were below and those who were above their traditional grade level in literacy took fewer quarters to complete their performance levels than did students at grade level, Westminster Public Schools, 2013/14



Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

However, more than 25 percent of these students completed their levels in three or fewer quarters. A larger percentage of students whose performance levels were above their traditional grade level completed their levels in three or fewer quarters than did students who were at grade level.

Teacher ratings of student competency had a small positive association with student academic achievement

Analyses were conducted to examine the relationship between teacher ratings of student competency and student academic achievement. Distributions of the learning target scores showed that teachers mostly assigned students learning target scores of 3.0 (see figure B1 in appendix B for an example). As a result, the distribution of performance-level competency scores, which aggregate each student's learning target scores, differed substantially from the distribution of student academic achievement as measured by state achievement test scores (see figures B2–B5 in appendix B). In particular, the distribution of student academic achievement resembled a normal distribution, while the distribution of performance-level competency scores did not.

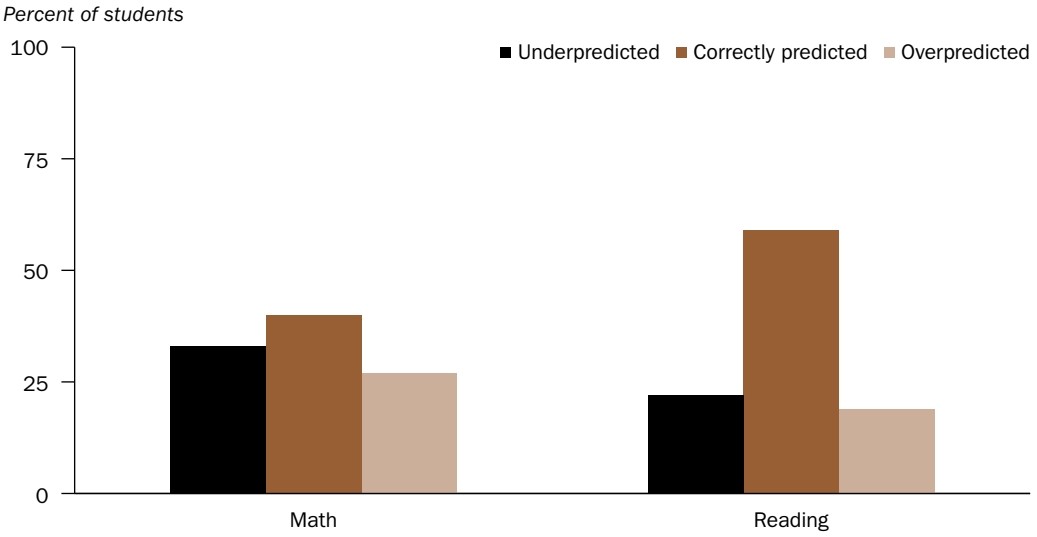
The performance-level competency scores accounted for only a small proportion (4 percent) of the variance in both math and reading scores

Performance-level competency scores had a statistically significant and positive relationship with state achievement test scores for both math and reading, but this relationship was weak. The performance-level competency scores accounted for only a small proportion (4 percent) of the variance in both math and reading scores (for more detailed findings, see appendix C). The relationships between performance-level competency scores and math and literacy test scores also did not vary significantly across students who were below, at, or above grade level.

Likewise, performance-level competency scores also had a significant and positive relationship with students' actual proficiency levels on state achievement tests, but this relationship was weak. When performance-level competency scores were used to predict student proficiency levels on the achievement tests, students' predicted proficiency levels corresponded to their actual proficiency levels for approximately 40 percent of students in math and 59 percent of students in reading (figure 3). A slightly greater percentage of students had predicted proficiency levels that were lower rather than higher than their actual proficiency levels. Again, only a small portion of the overall variance in students' actual proficiency levels was explained by students' proficiency level competency scores, indicating that the proficiency levels predicted by teachers' performance-level competency scores corresponded only to a limited degree to state achievement test scores.

Additional analyses were conducted to examine differences in the ability of performance-level competency scores to predict the proficiency levels of students who were below or above grade level. On average, the predicted proficiency levels of students below grade level were more likely to be higher than their actual proficiency level, while the predicted levels of students above grade level were more likely to be lower than their actual proficiency level (figure 4). Analyses conducted separately for students below or above grade level found these differences to be statistically significant. For example, among math students who were below grade level, predicted proficiency levels were higher than actual proficiency levels for 34 percent and lower than actual proficiency levels for 23 percent. Additionally, a substantially greater proportion of the predicted proficiency levels of students who were above grade level were lower than their actual levels (63 percent math and 42 percent literacy) compared with students below grade level (23 percent math and 13 percent literacy). For additional detail about these analyses, see appendix C.

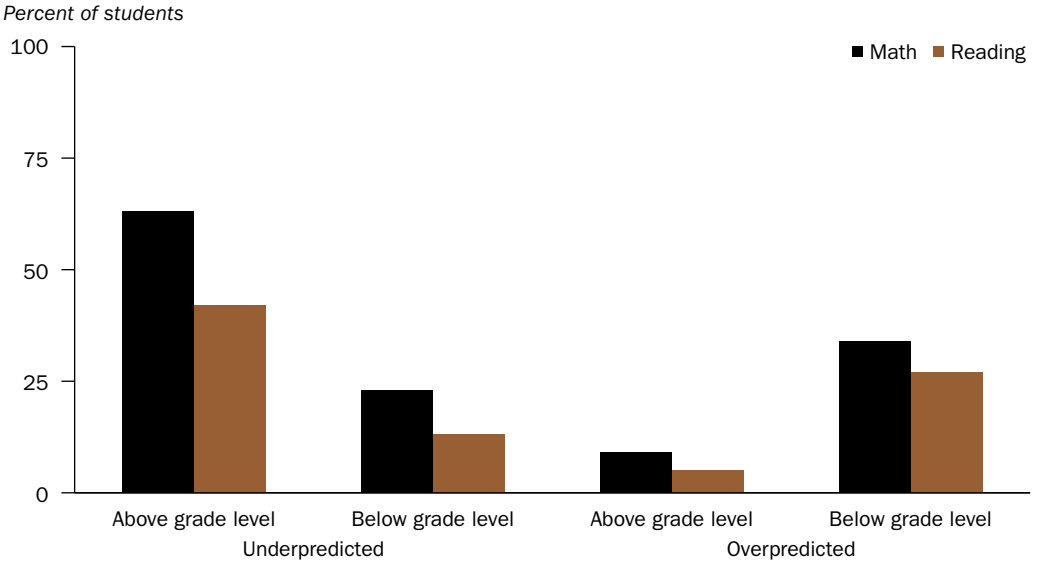
Figure 3. Performance-level competency scores correctly predicted the actual proficiency levels of 40 percent of students in math and 59 percent of students in reading, Westminster Public Schools, 2013/14



Note: Underpredicted refers to students whose predicted proficiency level was below their actual proficiency level, correctly predicted refers to students whose predicted proficiency level matched their actual proficiency level, and overpredicted refers to students whose predicted proficiency level was above their actual proficiency level.

Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

Figure 4. Predicted proficiency levels were more likely to be higher than actual proficiency levels for students below grade level, compared to students above grade level, Westminster Public Schools, 2013/14



Note: The state achievement test is the Transitional Colorado Assessment Program. Underpredicted refers to students whose predicted proficiency level was below their actual proficiency level, correctly predicted refers to students whose predicted proficiency level matched their actual proficiency level, and overpredicted refers to students whose predicted proficiency level was above their actual proficiency level.

Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

Implications of the study findings and limitations of the study

A majority of Westminster Public Schools students completed their courses of study in approximately four academic quarters. Although a majority of students who were in performance levels below or above their traditional grade level also completed their performance level in four quarters, a higher percentage of them than of students at grade level took three or fewer quarters. These results suggest that some students take advantage of the opportunity to complete performance levels in shorter periods of time—particularly those who are below or above grade level. Students below traditional grade level who complete performance levels more quickly may be benefiting from additional resources, such as more targeted instruction. However, it is also possible that teachers may have promoted students too quickly by scoring students as competent on their learning targets before they truly were competent. In fact, this is suggested by the finding that teachers' judgments of students' knowledge caused them to overpredict the actual proficiency levels of students below grade level; for these students, their actual proficiency level was below what would be expected given teachers' scores of their competency.

The results suggest that some students take advantage of the opportunity to complete performance levels in shorter periods of time—particularly those who are below or above grade level

This study found weaker relationships between teachers' judgments of students' knowledge and student achievement scores than other studies have found. On average, performance-level competency scores accounted for only about 4 percent of the variance in student achievement scores. Additionally, students' performance-level competency scores were found to accurately predict proficiency levels on the state achievement tests for 40 percent of students in math and 59 percent of students in reading. These results are smaller than those found in traditional schools: a prior study found that in samples of traditional schools, teachers' judgments of students' knowledge account for approximately 40 percent of variation in student achievement, while another study found teacher judgments of students' knowledge to be less than 50 percent accurate on average (Herman & Choi, 2008; Sudkamp et al., 2012). These findings are inconsistent with the theoretical hypotheses put forward by some researchers who conjectured that these relationships would be stronger in competency-based education (Marzano & Kendall, 2008).

There are several possible explanations for these unexpectedly smaller relationships. First, teachers of a given performance level had students of multiple traditional grade levels who took different grade-level state achievement tests. While the analyses described in this report transformed students' performance-level competency and state achievement test scores so that they were comparable within each performance level (this process is described in detail in appendix B), the Colorado Academic Standards addressed by the performance levels and by the Transitional Colorado Assessment Program do differ for students who are in performance levels different from their traditional grade levels. Although the state test is vertically scaled and consists of common items that are administered across grade levels so that scores are comparable across grade levels, the differences in the standards addressed in the performance level and the state achievement tests for students who are in a performance level that does not match their traditional grade level may be a cause of the loose relationship observed. However, analyses conducted only with students who were at grade level found similar results.

Second, the distributions of performance-level competency scores and student achievement test scores differed. The performance-level competency scores were not evenly distributed because most students received learning target scores of 3.0. The purpose of

the learning target scores that compose the performance-level competency scores was to indicate whether students achieved proficiency on the learning targets; the majority of students had learning target scores of 3.0 when the learning target data were collected for this study, at the time of the Transitional Colorado Assessment Program administration. The Transitional Colorado Assessment Program, in contrast, is a summative assessment designed to maximize variation, differentiate among students, have a normal distribution, and be used for accountability purposes. The lack of variability in the performance-level competency scores relative to the state achievement tests may weaken the relationship between the variables, as the performance-level competency scores do not differentiate students to the degree that the state achievement tests do. However, other results support the interpretation of a true weak association between teacher ratings and state achievement test performance. Performance-level competency scores that were transformed to have a similar distribution to the Transitional Colorado Assessment Program scores (see tables B1 and B2 in appendix B) also showed a weak relationship between the predicted and actual proficiency levels on the state achievement tests.

Third, there is evidence that teachers may not have used the Westminster Public Schools learning management system as intended. In many cases, students received the same score on all learning targets in a given performance level on the same date. Examination of the data and discussion with district leaders suggested that these cases likely corresponded with district testing (for example, benchmark assessments). This suggests that when students passed a district assessment, teachers entered into the learning management system scores designating students as proficient for all learning targets, rather than entering scores on an ongoing basis as they gathered proficiency information relevant for each individual learning target. In other cases, while some individual students' scores on the multiple learning targets within a particular performance level may have varied, they were provided on the same date, which roughly corresponded with student reporting deadlines. These findings may call into question the quality of the data entered into the district's learning management system. The examples of teachers' possible misapplication of the learning management system and the current study findings suggest that at least some teachers may not be using the Westminster Public Schools learning management system to track students' learning progression as intended. Rather, they may be entering data only as an administrative requirement. It is possible that teachers are using classroom assessment data to inform their instruction as intended by the district but are not entering the data into the learning management system, which was the primary source of data for this study.

A potential cause for the weak relationship between students' learning targets and state achievement test scores may be the result of recent changes in state academic standards and related district learning targets

Westminster Public Schools leaders noted that another potential cause for the weak relationship between students' learning targets and state achievement test scores may be the result of recent changes in state academic standards and related district learning targets. These changes may have led to a lack of clarity and fluency in teachers' use of the learning targets to assess student competency. Additionally, this study is limited in that it only used data entered into the Westminster Public Schools learning management system. It was beyond the scope of this study to examine teachers' actual assessment practices or the extent to which they utilized the learning management system as intended. District leaders may use the results of this study to inform professional development or policy changes regarding classroom assessment and use of learning management systems.

This study was conducted specifically within the context of Westminster Public Schools. Results may differ in schools and districts that use a different competency-based education

model. The descriptive and correlational nature of this study does not allow for causal attributions. For example, it is not possible to determine whether competency-based education helped students improve their learning or other academic outcomes. Future studies should be designed to examine the causal impact of competency-based education. Additional studies might also compare the progression of students in competency-based education contexts that differ from those in Westminster Public Schools. Studies could also examine how classroom implementation of competency-based education practices influences both student progression and teachers' judgments of student competency. This examination could include studying the influence of the clarity and sequence of learning targets on student academic outcomes and the influence of classroom assessment and instructional practices on student outcomes.

Appendix A. Westminster Public Schools learning target competency scale

Westminster Public Schools teachers used a scale adapted from the Marzano Taxonomy of Educational Objectives (Marzano & Kendall, 2008) to rate student competency of the learning targets within each performance level (table A1).

Table A1. Learning target scoring scale, Westminster Public Schools

Score	What the learner knows
4.0	In addition to Score 3.0, in-depth inferences and applications that go beyond what was taught.
3.5	In addition to Score 3.0 performance, in-depth inferences and applications with partial success.
3.0	No major errors or omissions regarding any of the information and processes (simple or complex) that were explicitly taught. (Learning Target)
2.5	No major errors or omissions regarding the simpler details and processes (Score 2.0 content) and partial knowledge of the more complex ideas and processes (Score 3.0 content).
2.0	No major errors or omissions regarding the simpler details and processes, but major errors or omissions regarding the more complex ideas and processes (Score 3.0 content). (Performance Indicators)
1.5	Partial understanding of the simpler details and processes (Score 2.0 content), but major errors or omissions regarding the more complex ideas and processes (Score 3.0 content).
1.0	With help, a partial understanding of some of the simpler details and processes (Score 2.0 content) and some of the more complex ideas and processes (Score 3.0 content).
0.5	With help, a partial understanding of some of the simpler details and processes (Score 2.0 content), but not the more complex ideas and processes (Score 3.0 content).
0.0	Even with help, no understanding or skill demonstrated.

Source: Westminster Public Schools.

Appendix B. Data and methodology

This appendix describes the study sample and details the data sources and study methodology. In the following discussion, performance-level completion refers to the sample, data, and methods used to address the first research question on how long it takes students who are below, at, or above grade level take to complete Westminster Public Schools math and literacy performance levels 3–8. Competency congruence refers to the sample, data, and methods used to address the second research question on the degree to which teachers' ratings of student competency relate to an external measure of student knowledge, the Transitional Colorado Assessment Program (TCAP).

Study sample

All data used in this report were provided by Westminster Public Schools. The district has a student enrollment of approximately 10,000 that is 48 percent female, 73 percent Hispanic, 41 percent English learner students, and 81 percent eligible for the federal school lunch program. The initial student sample consisted of all elementary and middle school students who were in a Westminster Public Schools math or literacy performance level 3–8 during the 2013/14 school year. Because data were not available for high school students, sample sizes are smaller for the higher performance levels. For example, high school students in a level 8 performance level are not represented, so the level 8 performance level includes only students who are at or below their traditional performance level. In contrast, the lower performance levels include students who are in performance levels above, at, and below their traditional grade levels. Because students are able to begin and complete performance levels at different speeds, some students may be in more than one performance level during a given school year. The study samples were defined by obtaining separate datasets from the Westminster Public Schools performance management system for each content area and performance level. The performance management system contains data ranging from 2010 to the present. Data gathered from the system for this study included students' performance level start and completion dates, learning target scores, and 2014 Transitional Colorado Assessment Program scores. To define the study samples used in this report, researchers restricted the samples based on students' performance level beginning and completion dates. Students were counted at a specific performance level only if administrative records showed they were in that level for 11 or more instructional days. Per district guidance, when new students are placed within a performance level, school staff have 11 instructional days to determine if the placements are appropriate or to recommend the students be placed in a different performance level.

All students were included regardless of when they started their performance level, such that some students started their level prior to the 2013/14 school year. However, per district guidance, students were excluded if they had been in a given performance level for more than two years (specifically, 346 instructional days). Westminster Public Schools leadership stated that students in a performance level for more than two years should receive intervention or otherwise no longer appear in the database; in these cases, students were likely students with special needs, students whose performance level completion date had not been entered into the system, or students who had left the district. District leadership recommended that students be included in the datasets only if they had been in their performance level for two years or fewer; students in the datasets for longer likely represent system error or students not representative of the general Westminster Public Schools population.

The nature of the data available through the Westminster Public Schools learning management system makes it difficult to define comparative samples. For example, at any given point in time, all students will appear in the system as having completed one performance level and not having completed another performance level, making it difficult to estimate how many students had not completed at least one performance level during a given period of time. While different criteria were used to define the various study samples, student composition across the samples was similar. The two samples of math students used to address research questions 1 and 2 both had 41 percent of students in a performance level at grade level. For the two literacy samples, 54 percent and 51 percent of students were at grade level. Across all samples, fewer than 10 percent of students were in a performance level above their traditional grade level.

Performance-level completion sample. The analysis sample included students who completed a math (N = 2,104) or literacy (N = 2,086) performance level at some time during the 2013/14 school year. Specifically, students were only included if they completed their performance level between July 1, 2013 and June 30, 2014, a period that includes all students who completed a performance level after the end of the 2013 summer session and those who completed a performance level during the 2014 summer session. (Students completing a performance level during a summer session are counted as having completed it during the previous academic year.)

Competency congruence sample. The analysis sample included students who were in math (N = 2,388) or literacy (N = 1,702) performance levels 3–8 during the 2013/14 school year and who had taken the 2014 math or reading TCAP assessment.

Data sources

Regional Educational Laboratory (REL) Central worked directly with Westminster Public Schools and the vendor of the district's learning management system to obtain data regarding the number of days students were in a Westminster Public Schools performance level and teacher ratings of student competency on the learning targets associated with each performance level and content area.

Performance-level completion data. The Westminster Public Schools learning management system allowed researchers to create datasets showing the date students started and completed their performance levels. Separate datasets were created for each content area (math and literacy) and performance level (levels 3–8). The number of instructional days students were in a performance level was calculated by determining the number of days between the day students began and completed a performance level and removing non-instruction days such as weekends, holidays, and teacher in-service days. In calculating the number of instructional days students were in a performance level, days spent in the level prior to the 2013/14 school year and during the 2014 summer session were included.

Competency congruence data. The Westminster Public Schools learning management system also allowed researchers to create datasets showing the learning target scores provided by teachers that students had received on specified dates (for the scale teachers used in providing learning target scores, see appendix A). Separate datasets were created showing the learning target scores students had for Westminster Public Schools math and

literacy performance levels 3–8 on March 28, 2014, the date recommended by the Colorado Department of Education for TCAP administration to be completed.

Westminster Public Schools provided REL Central with students' 2014 TCAP scaled scores and proficiency levels. Students are assessed annually in grades 3–8 in reading, writing, and math. In this study, only TCAP math and reading scores were used. The TCAP is administered to students at their traditional grade level in a paper-and-pencil format and is aligned with Colorado Academic Standards. Colorado Academic Standards and the TCAP are vertically aligned; the TCAP provides scale scores vertically aligned across the grades so there is continuity in scores across grade levels. Scores are expected to increase as grade level increases. To support this vertical alignment, the TCAP assessment contains common items that are administered across grade levels. Within each content area and grade level, students receive a TCAP scale score. These scale scores can be used to divide student performance into four performance levels using cutscores developed by the Colorado Department of Education: unsatisfactory, partially proficient, proficient, and advanced.

Creation of performance-level competency scores

Given that the Westminster Public Schools learning target scoring scale is not a true interval scale, an average of learning target scores within a performance level may not provide the most precise estimate of student competency or allow for efficient use in subsequent analyses. Furthermore, learning targets within a performance level vary in terms of how difficult they are to master. For example, the learning targets addressed near the end of a course are more difficult than those at the beginning. To provide a true interval scale estimate of teachers' rating of student competency within each content area and performance level, an overall performance-level competency scale score was created for each student using item response theory (IRT) methods.

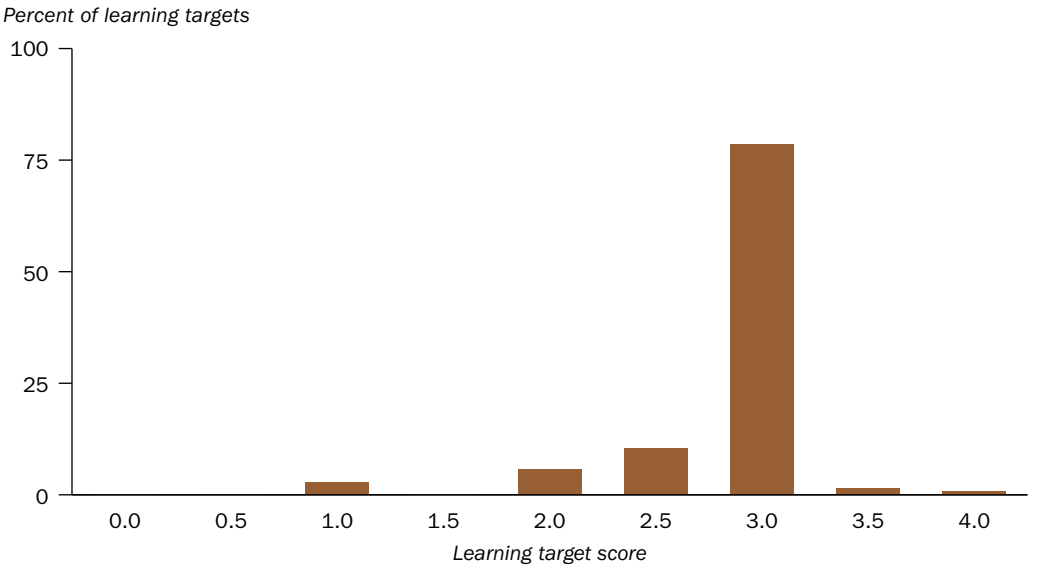
IRT analysis estimates the unique difficulty of each test item (in this case, each learning target) and uses this information to estimate student performance on the set of items (in this case, all learning targets within a performance level). Scores based on IRT analysis use all the information regarding the difficulty of each item to arrive at a student's score, so students get more credit for correctly answering difficult items. Accounting for learning target difficulty in creating an overall score of performance-level competency means student scores will more accurately and precisely represent student ability. Accurate and precise overall scores of student performance-level competency are needed to conduct the analyses required to address the second research question.

IRT analysis calibrates items by placing the items (based on difficulty) and the students (based on performance) on a scale whereby the increments of measurement are equal all along the scale. To create overall IRT scores, referred to in this report as performance-level competency scores, students' learning target scores were calibrated within each content area and performance level (for example, math 5) using the Andrich rating scale model. This scaling model is used for polytomous items that all use the same scale, such as questionnaire items using the same response scale (for example, 1 = Never, 3 = Sometimes, and 5 = Often). All Westminster Public Schools learning targets are rated by teachers using the Westminster Public Schools scoring scale (see appendix A). The Andrich rating scale model has the benefit of using all the information across all items and students during calibration to reduce the need for large numbers of students in each score level. Given that

the Westminster Public Schools scoring scale has eight score levels (in addition to 0), the Andrich rating scale is beneficial as it is likely that some levels have relatively few students. An example learning target score distribution for mathematics performance level 5 shows that, aggregated across students and the 35 math level 5 learning targets, the majority of learning targets (79 percent) have a score of 3 (figure B1). On average, approximately 36 percent of students did not have a learning target score for any given learning target.

Westminster Public Schools teachers provide learning target ratings if relevant information is available. For example, teachers may not provide ratings for learning targets that are addressed later in the semester until they have covered the relevant content. Because students may progress through course content at their own pace and start a given performance level at different times, the number of learning targets on which students have scores at a given point can vary; many students may not have any scores on several learning targets at a specific time. In assessment terms, these learning targets can be considered not administered, similar to items on an assessment that were administered to some students, but not to others, such as when students take different forms of the same assessment. Not administered items are treated as missing and allow students with incomplete data to remain in the sample. Students with missing learning target scores are not penalized by this approach; missing scores are not counted as poor performance on the learning target. IRT analysis uses all the existing data to estimate item difficulty and student ability, even among those students with incomplete data. To help ensure that Westminster Public Schools students had enough data, performance-level competency scores were only calculated for students who had scores for at least five learning targets within a performance level; this meant that approximately 13 percent of math students and 19 percent of literacy students were excluded from the initial study sample (see box 2). Omitting students with fewer than five learning target scores resulted in a change from 22 percent (whole sample) to 33 percent (restricted sample) of students who were behind grade level among math students, while the corresponding change was from 30 percent to 26 percent among literacy students.

Figure B1. An example learning target score distribution for math performance level 5, Westminster Public Schools, 2013/14



Source: Authors' analysis based on data provided by Westminster Public Schools; see text for details.

As discussed above, each performance level is associated with a defined number of common learning targets. Students must score proficient on all the learning targets before they may be promoted to the next level. At any given point, students may have scores on some learning targets and not others, depending on how long students have been in a performance level and what assessments have been administered. IRT was used to combine all the students' available learning target scores into an aggregated measure of proficiency within the performance level. Additionally, learning target scores are updated as teachers obtain additional information regarding student learning.

IRT analysis was conducted using the most current version (3.81.0) of Winsteps. Calibration places all the learning targets and all the students within a performance level onto a single, shared IRT scale where the learning target measures represent difficulty and the student performance-level competency scores represent ability. The logit scale created in the calibration is sample-dependent; so, although the logit scales within each performance level have a mean of 0 and a standard deviation of 1, this mean is a local mean based on the specific learning targets and sample used in the calibration. In other words, the 0/1 logic metric is not comparable among different grades and performance levels, meaning that the performance-level competency scores from Westminster Public Schools cannot be aggregated across performance levels without transformation. Following the IRT analysis, students' performance-level competency scores were outputted for subsequent analyses to address the second research question.

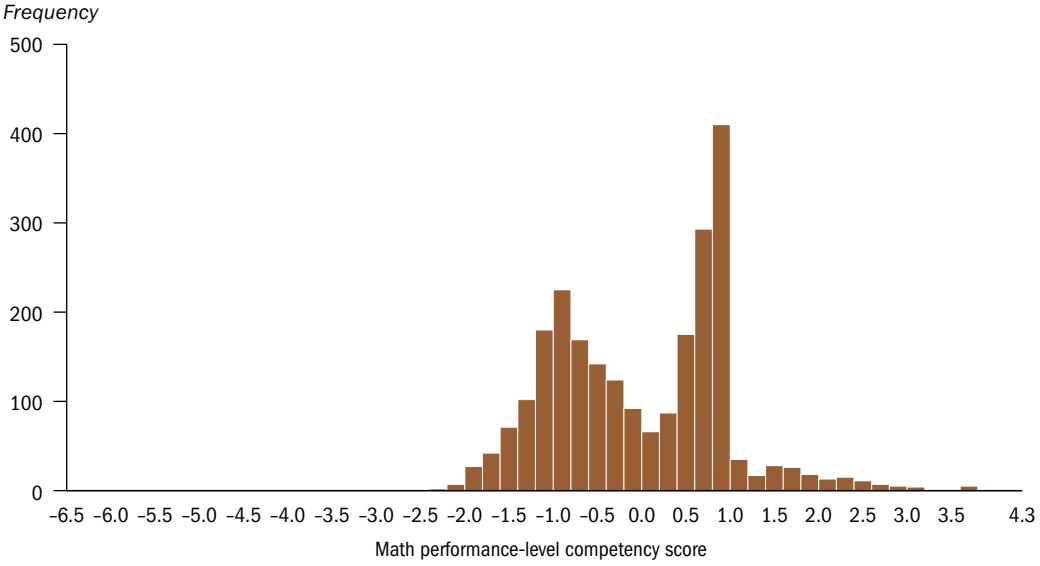
Westminster Public Schools literacy learning targets are divided into a Reading for All Purposes strand and a Writing and Composition strand. For this report, literacy performance-level competency scores were computed using only the learning targets in the Reading for All Purposes strand. All math learning target data were used to calculate math performance-level competency scores. The resulting math and literacy performance-level competency scores were used to address the second research question. The process of aggregating all of each student's learning target scores resulted in student-level performance-level competency scores that were on an interval scale with a greater distribution of scores than those of the individual learning targets. However, the performance-level competency scores did have a non-normal distribution (figures B2 and B3). These are compared with the normal distributions seen among the TCAP scores (figures B4 and B5).

A sensitivity analysis was conducted to determine whether study results associated with research question two are sensitive to the IRT transformation. Hierarchical linear modeling analyses were conducted using the number of learning target scores students had completed (earning a score of 3 or higher) as the independent variable. Students with fewer than five learning target scores were included in the analyses. Results showed a significant relationship with students' math TCAP scores ($B = .046$, $p = .027$) and a nonsignificant relationship with students' reading TCAP scores ($B = .018$, $p = .405$). Results of the sensitivity analysis indicated a weak relationship between the learning target scores and student achievement. Analyses conducted with the IRT transformed data showed stronger relationships between the variables (see appendix C).

Data analysis

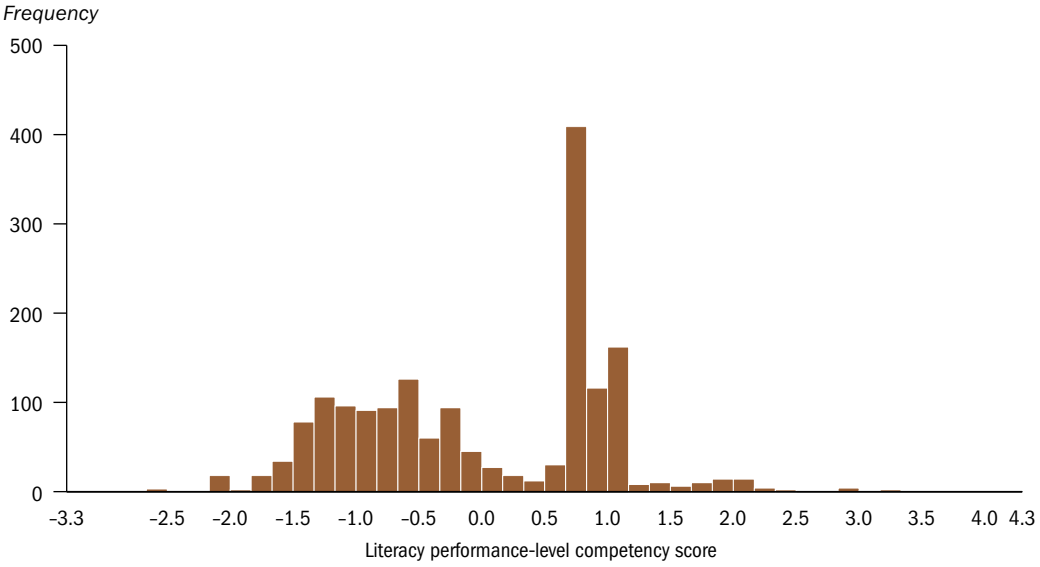
This section describes the analytic approaches used to address research questions one and two.

Figure B2. Math performance-level competency score distribution, Westminster Public Schools, 2013/14



Source: Authors' analysis based on data provided by Westminster Public Schools; see text for details.

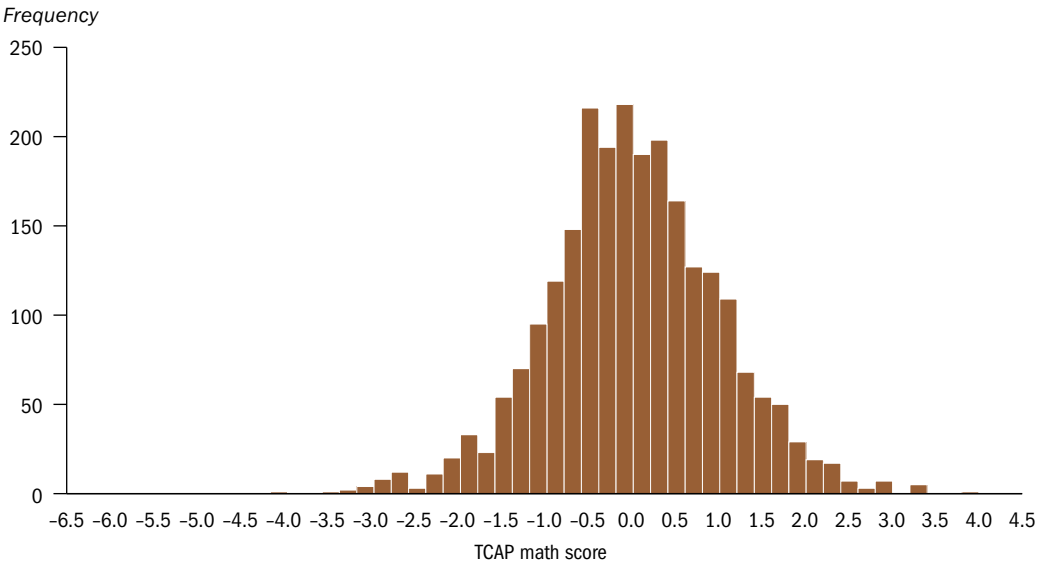
Figure B3. Literacy performance-level competency-score distribution, Westminster Public Schools, 2013/14



Source: Authors' analysis based on data provided by Westminster Public Schools; see text for details.

Performance-level completion analyses. To answer the first research question, the number of instructional days each student took to complete math and literacy performance levels 3–8 was calculated. Any days a student was in a performance level prior to the 2013/14 school year were included in the calculations. Within each content area and across performance levels the average number of instructional days students were in a performance level was separately calculated for students at each level below grade level, students at grade level, and students at each level above grade level. For example, the average number

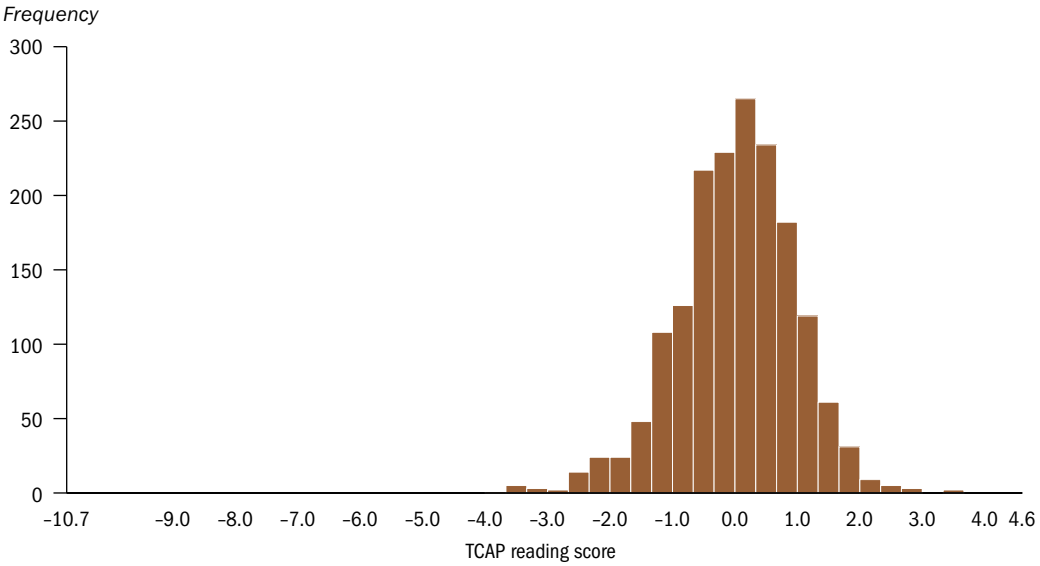
Figure B4. Score distribution on the Transitional Colorado Assessment Program math achievement test, Westminster Public Schools, 2013/14



TCAP is Transitional Colorado Assessment Program.

Source: Authors' analysis based on data provided by Westminster Public Schools; see text for details.

Figure B5. Score distribution on the Transitional Colorado Assessment Program reading achievement test, Westminster Public Schools, 2013/14



TCAP is Transitional Colorado Assessment Program.

Source: Authors' analysis based on data provided by Westminster Public Schools; see text for details.

of instructional days was calculated for all students who were in a performance level two levels below their traditional grade level. Students were designated below, at, or above grade level by comparing their traditional grade level during the 2013/14 school year with their given Westminster Public Schools performance level.

The number of instructional days students were in a performance level was used to categorize the number of academic quarters students took to complete their performance level. During the 2013/14 school year, the district's academic quarters consisted together of 173 instructional days. Estimated quarters were calculated based on the average of 42 instructional days per quarter.

Students in Westminster Public Schools begin and complete performance levels at various times. The calculation of the number of academic quarters to complete a performance level was based on the number of instructional days during the 2013/14 school year. These numbers were used to estimate the quarters to complete for all students, regardless of their performance level beginning and ending dates. These estimates may not represent the actual 2013/14 academic quarter in which a performance level was completed, but they do provide a general estimate of the number of quarters it took to complete a performance level.

Competency congruence analyses. Two methods were used to address the second research question. First, hierarchical linear modeling was used to estimate the correlation between students' performance-level competency scores and TCAP scores. Separate analyses were conducted for math and reading, aggregating across performance levels, by first standardizing students' performance-level competency scores and TCAP scores within each performance level. Because students can progress at their own pace, at the time of TCAP testing (which is administered to students according to traditional grade status), students may not have been in the corresponding Westminster Public Schools performance level that they were tested in. For this reason and because students may have multiple teachers during a school year, students were nested in schools rather than by teachers in all analyses. While students take the TCAP at their traditional grade level, the TCAP is vertically scaled so that there is continuity in scores across grade levels: scores are expected to increase as grade level increases. This scaling allows for the examination of the relationship between student performance-level competency scores within a given course of study to the TCAP scores of students of multiple traditional grade levels within that course of study.

Within each Westminster Public Schools content area, students' TCAP scores were regressed on their IRT performance-level competency scores, including students of all traditional grade levels within the given Westminster Public Schools content and performance levels. As described above, TCAP provides vertically scaled scores, so scores from all grade levels are on a common scale. In all analyses, students were nested in schools. Because students can progress through performance levels at varying rates and times, with resulting different teachers during the school year, REL Central researchers did not believe that nesting students within the teacher of record at the time of TCAP testing was appropriate. The model used to assess the relationship between the two variables is as follows:

Level 1: $Y_{ij} = \beta_{0j} + \beta_{1j}(\textit{performance-level competency score})_{ij} + e_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

$\beta_{1j} = \gamma_{10}$

where Y_{ij} is the TCAP scale score for student i in school j; β_{0j} is the school mean (intercept) for school j; β_{1j} is the within-school association between performance-level competency

scores and TCAP scores; $(performance\text{-}level\ competency\ score)_{ij}$ is the performance-level competency score for student *i* in school *j*; e_{ij} is the student-level residual for student *i* in school *j*; γ_{00} is the grand (overall average) mean; u_{0j} is the school-level residual for school *j*; *and* γ_{10} is the average (grand) association between performance-level competency scores and TCAP scores. Secondary analyses examined the extent to which the relationship between performance-level competency scores and TCAP achievement differed for students who were below, at, or above grade level. To assess this relationship, a variable (at-level) representing the number of performance levels students were away from their traditional grades was entered into the equation, along with an at-level by performance-level competency score interaction term. This modeling approach necessarily assumes the interaction effect is constant across the student levels.

To evaluate each of the models, the significance and size of the regression coefficient related to the performance-level competency score variable were examined. To understand the relationship between performance-level competency and student achievement, the variance components in the full model above were compared with the variance components of a null model in which the performance-level competency score was not included. The total variances in each model were used to estimate the amount of variance in student achievement scores accounted for by performance-level competency scores. The specific equation was (null model variance – full model variance)/null model variance. The at-level by performance-level competency score interaction was assessed by examining the significance of the interaction terms' regression coefficient. A statistically significant interaction would indicate that the relationship between performance-level competency scores and TCAP scores is different for students who are below, at, or above grade level.

To explore the relationship between performance-level competency scores and student academic achievement, students' performance-level competency scores were used to predict the students' TCAP proficiency levels, using the TCAP cutscore ranges associated with students' Westminster Public Schools performance level rather than their traditional grade level. Equipercentile equating was used to link students' performance-level competency scores for each performance level separately to the TCAP scale distribution for students of all traditional grade levels within each Westminster Public Schools content area and performance level. In this method, scores on the two tests at the same percentile are considered to indicate the same level of performance (Kolen & Brennan, 1995). The equipercentile equating was used to transform the score distributions on the logit scale from the IRT calibrations for each content area and performance level to the corresponding TCAP scale. This transformation retains the interval characteristic of the data provided by the IRT calibration and results in the assignment of each student to a predicted TCAP score. The distributions for students' math and reading TCAP and equipercentile-transformed classroom competency scores are presented in tables B1 and B2. The data in these tables show that the equipercentile transformed distributions for the performance-level competency scores were close to the TCAP distributions in terms of overall variance and mean level.

The predicted TCAP scores were then used to assign each student a predicted proficiency level using the 2014 TCAP proficiency level cutscore ranges for the grade that corresponded to each student's Westminster Public Schools performance level. Similarly, a student's original TCAP score was used to obtain an actual TCAP proficiency level using the cutscores that corresponded to the relevant Westminster Public Schools performance

Table B1. Score distributions for Transitional Colorado Assessment Program math achievement tests and transformed performance-level competency scores, Westminster Public Schools, 2013/14

Math performance level	*N*	Mean	Standard deviation	Skewness	Kurtosis
Math 3 TCAP	633	454.35	54.78	−0.007	0.343
Math 3 transformed	633	450.36	56.90	−0.173	0.362
Math 4 TCAP	747	487.35	52.58	−0.057	0.850
Math 4 transformed	747	485.38	54.44	−0.190	1.210
Math 5 TCAP	607	506.04	46.94	−0.144	0.740
Math 5 transformed	607	504.67	45.50	−0.126	1.393
Math 6 TCAP	245	547.18	49.90	0.145	−0.189
Math 6 transformed	245	537.86	48.90	0.196	−0.179
Math 7 TCAP	137	593.15	35.88	0.262	1.501
Math 7 transformed	137	560.46	50.63	0.767	0.320
Math 8 TCAP	35	606.40	27.92	0.573	−0.025
Math 8 transformed	35	613.32	55.51	3.889	18.525

TCAP is Transitional Colorado Assessment Program.

Source: Authors' analysis based on data provided by Westminster Public Schools; see text for details.

Table B2. Score distributions for Transitional Colorado Assessment Program reading achievement test and transformed performance-level competency scores, Westminster Public Schools, 2013/14

Reading performance level	*N*	Mean	Standard deviation	Skewness	Kurtosis
Reading 3 TCAP	413	559.92	44.35	−1.706	13.952
Reading 3 transformed	413	556.29	42.82	−0.564	1.819
Reading 4 TCAP	492	585.88	39.58	−1.435	9.502
Reading 4 transformed	492	582.17	40.64	−1.442	8.665
Reading 5 TCAP	351	606.95	38.27	−1.359	7.112
Reading 5 transformed	351	603.20	27.21	−0.480	0.443
Reading 6 TCAP	175	617.20	45.51	−0.417	−0.019
Reading 6 transformed	175	617.19	41.46	−0.214	0.078
Reading 7 TCAP	190	637.85	38.40	−0.376	0.592
Reading 7 transformed	190	636.28	37.41	−0.136	0.196
Reading 8 TCAP	95	646.33	42.61	−0.571	1.409
Reading 8 transformed	95	646.33	35.18	0.299	−0.234

TCAP is Transitional Colorado Assessment Program.

Source: Authors' analysis based on data provided by Westminster Public Schools; see text for details.

level. For example, the grade 5 TCAP cutscores were used for all students in Westminster Public Schools performance level 5 to determine their predicted and actual TCAP proficiency levels, regardless of their traditional grade level. Comparisons of the distributions in students' actual (not predicted) proficiency levels using the TCAP cutscores associated with their traditional grade levels versus their performance levels show that the distributions are relatively similar; a greater proportion of students score proficient and advanced in the transformed distributions (table B3). For example, 59 percent versus 74 percent of

literacy students were classified as proficient or advanced by using the cutscores associated with their traditional grades versus their performance levels. For this reason, a sensitivity analysis was conducted to examine the correlation between students' actual and predicted TCAP proficiency levels among only students who were at grade level (in other words, those students whose traditional grade and performance level cutscores were the same). Spearman's rank correlation coefficients were nearly identical to those found with the larger sample (r_s = .126, p < .000) and reading (r_s = .101, p < .01; see table C6 in appendix C for the results for the whole student sample).

The agreement between students' predicted and actual performance level–adjusted proficiency levels was examined by computing cross-classification tables for each content area and performance level. These tables (see tables C7 and C8 in appendix C) show the number and percent of students who had the same predicted and actual proficiency levels, as well as the number and percent of students whose predicted and actual proficiency levels did not correspond. The Spearman rank correlation coefficient (r_s) was also calculated to provide an index of correspondence between the predicted and actual performance levels.

Table B3. Proficiency level distributions using cutscores based on students' traditional grade levels versus Westminster Public Schools performance levels, Westminster Public Schools, 2013/14 (percent)

Proficiency level content area	Unsatisfactory	Partially proficient	Proficient	Advanced
Original TCAP, Math	18.4 (*n* = 444)	32.1 (*n* = 775)	36.7 (*n* = 886)	12.6 (*n* = 304)
Transformed TCAP, Math	2.7 (*n* = 66)	24.7 (*n* = 596)	56.6 (*n* = 1,367)	15.7 (*n* = 380)
Original TCAP, Reading	8.8 (*n* = 173)	31.5 (*n* = 617)	57.6 (*n* = 1,129)	1.3 (*n* = 25)
Transformed TCAP, Reading	3.6 (*n* = 71)	22.0 (*n* = 431)	72.4 (*n* = 1,420)	1.1 (*n* = 22)

TCAP is Transitional Colorado Assessment Program.

Source: Authors' analysis based on data provided by Westminster Public Schools; see text for details.

Appendix C. Additional study findings

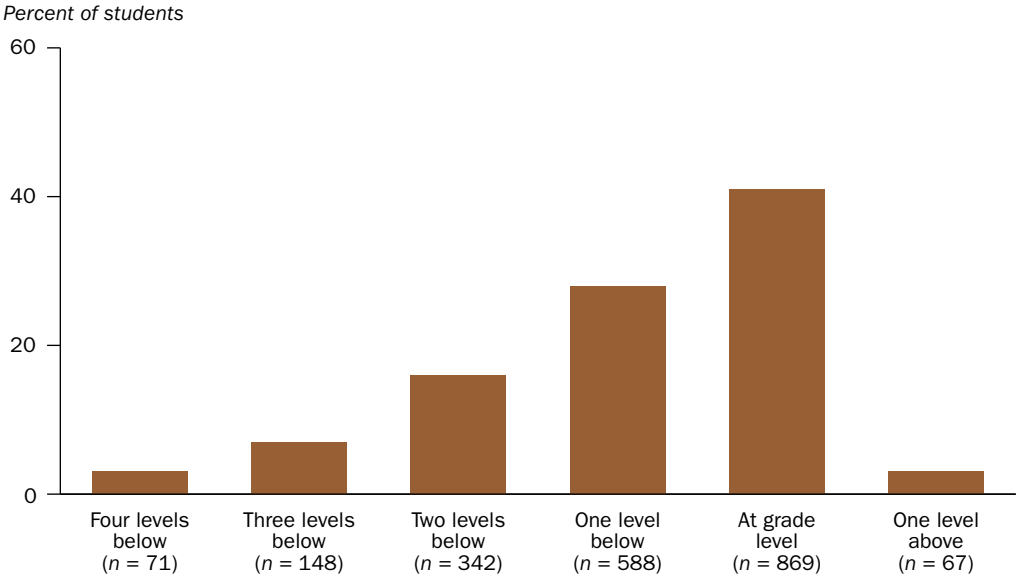
This appendix provides additional detail on the study findings. In the following discussion, performance-level completion refers to the sample, data, and methods used to address the first research question (How long do students who are below, at, and above grade level take to complete Westminster Public Schools math and literacy performance levels 3–8?). Competency congruence refers to the sample, data, and methods used to address the second research question (To what degree do teachers' ratings of student competency relate to an external measure of student knowledge?).

Performance-level completion results

During the 2013/14 school year 88 percent and 94 percent of Westminster Public Schools students who completed a math or literacy performance level were between two levels below and one level above their traditional grade levels, respectively (figures C1 and C2). Fewer than 3 percent of students took more than five quarters to complete their level, although the sample omitted students who were in a performance level for more than two years.

In examining the amount of time students took to progress through Westminster Public Schools math and literacy performance levels 3–8, the average number of days and the number of quarters students took to complete each level were calculated. The average number of days students took to complete each math and literacy performance level are presented in table C1.

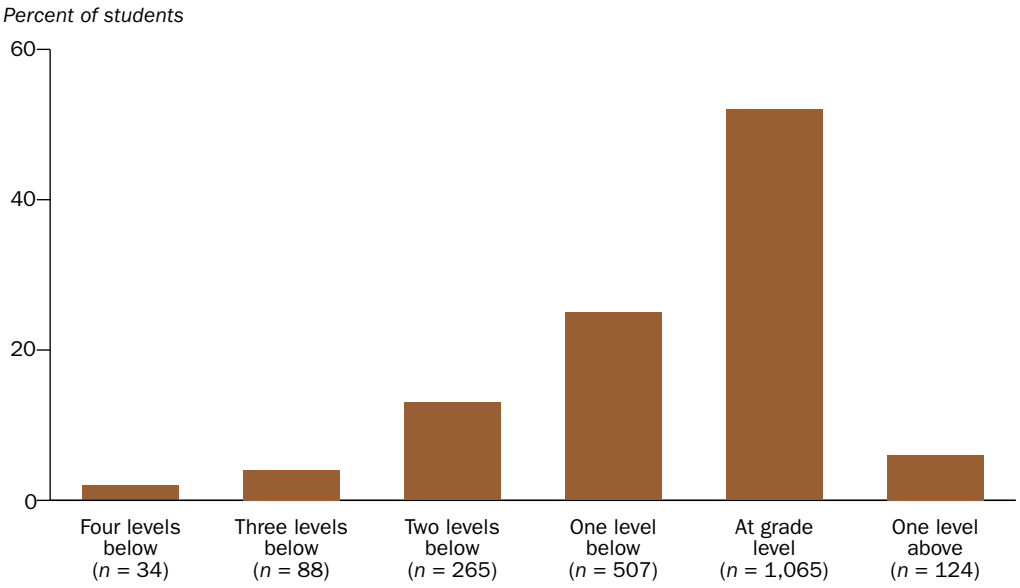
Figure C1. Percentage of math students in performance levels below, at, or above their traditional grade levels, Westminster Public Schools, 2013/14



Note: Fewer than 1 percent of Westminster Public Schools students were in a performance level five or more below or two or more above their traditional grade levels.

Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

Figure C2. Percentage of literacy students in performance levels below, at, or above their traditional grade levels, Westminster Public Schools, 2013/14



Note: Fewer than 1 percent of Westminster Public Schools students were in a performance level five or more below or two or more above their traditional grade levels. Numbers in parentheses represent sample sizes.

Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

In presenting the following results, data were aggregated across performance levels. On average, students took between 110 and 150 instructional days to complete a performance level (table C2).

The number of academic quarters students who were in a math or literacy performance level took to complete their levels was determined and percentages were calculated on the basis of the total number of students in each row (for example, all students in math performance levels 3–8 who were one performance level behind grade level; tables C3 and C4). For example, 77 percent of students who were in a math performance level corresponding to their traditional grade completed that level in four quarters, whereas 29 percent of students who were two levels behind their traditional grade completed their level in one quarter.

Table C1. Average number of days to complete performance levels, Westminster Public Schools, 2013/14

	Performance level					
Content area	3	4	5	6	7	8
Math	128.00 [51.29] (*n* = 551)	122.27 [58.13] (*n* = 626)	134.47 [52.38] (*n* = 554)	130.96 [54.37] (*n* = 212)	150.26 [40.56] (*n* = 113)	166.25 [20.40] (*n* = 48)
Literacy	132.35 [52.28] (*n* = 462)	137.40 [50.59] (*n* = 527)	125.10 [56.64] (*n* = 425)	130.68 [54.57] (*n* = 251)	139.66 [53.17] (*n* = 281)	153.93 [43.40] (*n* = 140)

Note: Numbers in brackets are standard deviations.

Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

Table C2. Average number of days students who were in a performance level below, at, or above grade level took to complete the performance level, Westminster Public Schools, 2013/14

Content area	2 or more below	2 below	1 below	At grade	1 above
Math	123.68 [59.25] (*n* = 238)	110.38 [63.98] (*n* = 342)	122.21 [55.25] (*n* = 588)	145.04 [41.38] (*n* = 869)	137.24 [52.33] (*n* = 67)
Literacy	119.39 [57.13] (*n* = 122)	117.06 [57.58] (*n* = 265)	117.31 [59.22] (*n* = 507)	150.21 [40.88] (*n* = 1065)	122.73 [61.95] (*n* = 124)

Note: Numbers in brackets are standard deviations.

Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

Table C3. Percentage of students completing a math performance level by number of quarters, Westminster Public Schools, 2013/14

Number of quarters	Two below	One below	At grade	One above
One	28.65	17.86	5.18	14.93
Two	6.73	5.95	6.44	2.99
Three	11.70	21.60	9.90	10.45
Four	52.34	53.23	77.10	70.15
Five	0.29	0.85	1.15	1.49
More than 5	0.29	0.51	0.23	0.00
Number of students	342	588	869	67

Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

Table C4. Percentage of students completing a literacy performance level by number of quarters, Westminster Public Schools, 2013/14

Number of quarters	Two below	One below	At grade	One above
One	18.49	23.27	6.57	25.00
Two	12.45	8.48	3.47	4.03
Three	13.58	11.24	6.67	4.84
Four	54.34	56.21	74.65	57.26
Five	1.13	0.59	8.45	8.87
More than 5	0.00	0.20	0.18	0.00
Number of students	265	507	1,065	124

Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

Competency congruence results

To examine the relationship between students' performance-level competency and achievement scores, hierarchical linear regression was first used to estimate the correlation between the measures, accounting for students nested in schools. The standardized regression coefficient, significance levels, interclass correlation coefficients, and the proportion of variance in the dependent variable (student academic achievement) accounted for by the dependent variable (performance-level competency) were calculated (table C5).

Table C5. Hierarchical linear modeling results examining the relationships between students' class competency and achievement scores, Westminster Public Schools, 2013/14

Independent variable	Standardized β	*p* value	Intraclass correlation coefficient	Explained variance
Math competency	.203	.000	.103	.04
Math competency x at-level interaction	.020	.237	—	—
Grade-level sensitivity analysis	.255	.000	.111	.07
Literacy competency	.208	.000	.074	.03
Literacy competency x at-level interaction	–.010	.820	—	—
Grade-level sensitivity analysis	.251	.000	.074	.06

Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

Interaction terms were then entered in to the equations to determine whether the relationship between performance-level competency scores and student academic achievement differed for students who were below, at, or above their traditional grade levels. Finally, a sensitivity analysis was conducted that included only students who were in a performance level that corresponded to their traditional grade level.

To examine the relationship between teachers' ratings of student competency and student academic proficiency levels, students' performance-level competency scores were used to predict their Transitional Colorado Assessment Program (TCAP) proficiency levels, using the TCAP proficiency-level cutscore ranges associated with their Westminster Public Schools performance levels rather than their traditional grade levels. Spearman's rank correlation coefficients (r_s) indicated the level of agreement between students' predicted and actual performance level–adjusted proficiency levels for the aggregated sample of students and separately for students who were below, at, or above grade level (table C6).

Cross-classification results indicated the number and percentage of students in each content area and performance level whose predicted proficiency level corresponded to their actual performance level–adjusted proficiency level (tables C7 and C8).

Table C6. Correlations between students' predicted and actual performance level–adjusted proficiency levels for students below, at, or above grade level, Westminster Public Schools, 2013/14

	Math		Reading	
Amount above or below grade level	*N*	r_s	*n*	r_s
Overall	2,388	.125***	1,702	.103***
2 below	392	.135**	212	.171*
1 below	728	.208***	433	.185***
At grade level	968	.126***	870	.099**
1 above	54	.170	88	.112

*Significant at $p < .05$; ** significant at $p < .01$; *** significant at $p < .001$.

Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

Table C7. Math cross-classification table showing the degree of correspondence between students' predicted and actual performance level–adjusted proficiency level on the Transitional Colorado Assessment Program, Westminster Public Schools, 2013/14

Transitional Colorado Assessment Program proficiency level	Equipercentile predicted proficiency level									
	Unsatisfactory		Partially proficient		Proficient		Advanced		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Unsatisfactory	7	10.61	20	30.30	34	51.52	5	7.58	66	100.00
Partially proficient	20	3.37	187	31.53	334	56.32	52	8.77	593	100.00
Proficient	43	3.18	431	31.83	692	51.11	188	13.88	1,354	100.00
Advanced	2	0.53	72	19.20	231	61.60	70	18.67	375	100.00

Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

Table C8. Reading cross-classification table showing the degree of correspondence between students' predicted and actual performance level–adjusted proficiency level, Westminster Public Schools, 2013/14

Transitional Colorado Assessment Program proficiency level	Equipercentile predicted proficiency level									
	Unsatisfactory		Partially proficient		Proficient		Advanced		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Unsatisfactory	5	7.94	22	34.92	36	57.14	0	0.00	63	100.00
Partially proficient	18	4.70	116	30.29	247	64.49	2	0.52	383	100.00
Proficient	43	3.48	291	23.54	885	71.60	17	1.38	1,236	100.00
Advanced	0	0.00	4	20.00	15	75.00	1	5.00	20	100.00

Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

Cross-classification results also showed the percentage of students whose proficiency levels were underpredicted, correctly predicted, and overpredicted (table C9).

McNemar-Bowker Tests of Internal Symmetry were conducted separately for students in performance levels below, at, or above grade level to determine if the percentage of students

Table C9. Percentage of students below, at, or above grade level whose performance level–adjusted proficiency level was underpredicted, correctly predicted, or overpredicted, Westminster Public Schools, 2013/14

Amount above or below grade level	Math			Reading		
	Underpredicted	Correctly predicted	Overpredicted	Underpredicted	Correctly predicted	Overpredicted
Overall	33.46	40.03	26.51	21.80	59.17	19.04
2 below	21.43	42.09	36.48	11.79	59.43	28.77
1 below	24.59	43.41	32.01	14.09	60.51	25.40
At grade level	46.59	36.67	16.74	27.70	58.85	13.44
1 above	62.96	27.78	9.26	42.05	53.41	4.55

Note: Underpredicted refers to students whose predicted proficiency level was below their actual proficiency level, correctly predicted refers to students whose predicted proficiency level matched their actual proficiency level, and overpredicted refers to students whose predicted proficiency level was above their actual proficiency level.

Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

underpredicted and the percentage of students overpredicted were significantly different. Results showed these differences to be significant for the majority of comparisons (table C10).

Table C10. McNemar-Bowker test of differences in the percentage of students whose performance level–adjusted proficiency level was underpredicted versus overpredicted, Westminster Public Schools, 2013/14

Amount above or below grade level	Math		Reading	
	Chi square value	*p* value	Chi square value	*p* value
Overall	22.28	.001	5.41	.368
2 below	40.95	.000	16.91	.002
1 below	10.42	.064	23.68	.000
At grade level	143.00	.000	48.88	.000
1 above	22.32	.000	27.29	.000

Note: Underpredicted refers to students whose predicted proficiency level was below their actual proficiency level, and overpredicted refers to students whose predicted proficiency level was above their actual proficiency level.

Source: Authors' analysis based on data provided by Westminster Public Schools; see appendix B for details.

References

Adams County School District 50. (2011). *Standards-based scoring.* Westminster, CO: Author. Retrieved November 4, 2014, from http://wiki.adams50.org/mediawiki/images/5/57/Adams_County_School_District_50_SBS_Scoring.pdf.

Herman, J. L., & Choi, K. (2008). *Formative assessment and the improvement of middle school science learning: The role of teacher accuracy* (CRESST Report No. 740). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, University of California. http://eric.ed.gov/?id=ED502630

Kolen, M. J., & Brennan, R. L. (1995). *Test equating methods and practices.* New York, NY: Springer-Verlag.

Marzano, R. J. (2006). *Classroom assessment and grading that work.* Alexandria, VA: Association for Supervision and Curriculum Development. http://eric.ed.gov/?id=ED509063

Marzano, R. J., & Kendall, J. S. (2008). *Designing and assessing educational objectives: Applying the new taxonomy.* Thousand Oaks, CA: Corwin Press.

Patrick, S., & Sturgis, C. (2013). *Necessary for success: Building mastery of world-class skills: A state policymakers guide to competency education.* Vienna, VA: International Association for K–12 Online Learning. http://eric.ed.gov/?id=ED561282

Sturgis, C., & Patrick, S. (2010). *When success is the only option: Designing competency-based pathways for next generation learning.* Vienna, VA: International Association for K–12 Online Learning. http://eric.ed.gov/?id=ED514891

Sudkamp, A., Kaiser, J., & Moller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743–762. http://eric.ed.gov/?id=EJ993888

White, T. (2014). *Giving credit where credit's due: A 50-state scan of course credit policies.* Stanford, CA: Carnegie Foundation for the Advancement of Teaching. Retrieved November 1, 2015, from https://www.carnegiefoundation.org/wp-content/uploads/2013/08/CUP_Policy_MayUpdate1.pdf.

Wolfe, R. E. (2012). *Aligning competencies to rigorous standards for off-track youth: A case study of Boston Day and Evening Academy.* Boston, MA: Jobs for the Future. http://eric.ed.gov/?id=ED539877

The Regional Educational Laboratory Program produces 7 types of reports

Making Connections
Studies of correlational relationships

Making an Impact
Studies of cause and effect

What's Happening
Descriptions of policies, programs, implementation status, or data trends

What's Known
Summaries of previous research

Stated Briefly
Summaries of research findings for specific audiences

Applied Research Methods
Research methods for educational settings

Tools
Help for planning, gathering, analyzing, or reporting data or research