



STATEMENTS OF SCORE INTERPRETATIONS AND USES (SIUs) FOR THE NEW MEXICO MEASURES OF STUDENT SUCCESS AND ACHIEVEMENT (NM-MSSA) ASSESSMENT

Mathematics

Score Interpretation and Use (SIU) Statements for the NM-MSSA and NM-ASR Assessment Programs

The phrase “intended score interpretations for uses” appears several times in the *Standards for Educational and Psychological Testing* and is at the core of the field’s views on validity and validation. It also is central to responding successfully to USDE peer review requirements. For the NM-MSSA ELA and mathematics assessments, the phrase refers broadly to **test scores** (i.e., total test scale scores, subtest indicators), **aggregations of test scores** (e.g., the percentage of students at and above Level 3: Proficient), and **other test performance informational elements** (e.g., the definition of Proficient in the Proficiency Level Descriptors).

SIU Statements for the NM-MSSA: Mathematics

Using this broad interpretation of the phrase, the intended score interpretations and uses for NM-MSSA are stated below. These statements reflect input from multiple statewide webinars with educator and parent stakeholders and PED’s Technical Advisory Committee.

PED and Cognia will use the final, approved SIU statements to guide decisions about test design and score reporting.

Intended Interpretations and Uses for Individual Students and Groups of Students

Score Interpretation/Use Statement	Explanation/Annotation
NM-MSSA Program Purpose Statements	
<p align="center">Program Purpose Statement, Grades 3–8 NM-MSSA</p> <p>The NM-MSSA grades 3–8 assessments are designed to measure whether students are on track to be ready for college or career, as defined by the state, by showing they have mastered <i>the New Mexico Common Core State Standards</i> for mathematics. Results are presented using scale scores and proficiency levels.</p> <p>Proficient performance in each grade indicates both mastery of currently assessed grade level and preceding grades’ expectations and progress toward college and career readiness.</p>	<p>NM-MSSA scores should be interpreted in relation to the <i>New Mexico Common Core State Standards</i> that are targeted by the assessment.</p> <p>College readiness indicates that a student is prepared to enter directly into and succeed (i.e., earn a C or better) in entry-level, credit-bearing college and relevant technical courses at two- and four-year public institutions of higher education, without the need for remediation.</p>



Score Interpretation/Use Statement	Explanation/Annotation
<p>The mathematics standards require a solid understanding of concepts, a high degree of procedural skill and fluency, and the application of mathematics to solve problems. See details at https://webnew.ped.state.nm.us/bureaus/assessment-3/nm-mssa/.</p>	<p>Career readiness indicates that students have developed the academic and technical skills (i.e., workplace competencies in one or more of 16 career clusters) necessary to succeed in future careers and to become lifelong learners.</p> <p>College and Career Readiness is defined by the State and can be found in the following College and Career Readiness Bureau’s web page: https://webnew.ped.state.nm.us/bureau/college-career-readiness/</p>
<p>Program Purpose Statement, Grade 8 NM-MSSA</p> <p>Performance on the grade 8 NM-MSSA indicates mastery of (a) grade 8 content standards, and (b) solid understanding of concepts, a high degree of procedural skill and fluency, and the application of mathematics to solve problems. See details at https://webnew.ped.state.nm.us/bureaus/assessment-3/nm-mssa/.</p> <p>It also is (c) a predictor of being on track for college and career readiness as defined by the College Board’s <i>College and Career Readiness Benchmarks</i>: see https://collegereadiness.collegeboard.org/about/scores/benchmarks. Cognia will establish a psychometric link from grade 8 NM-MSSA scores to College Board PSAT scores that will enable monitoring of student progress toward the College Board CCR Benchmarks.</p>	<p>The program purpose statements apply in grades 3–8.</p> <p>In addition, performance on the grade 8 NM-MSSA can be interpreted as a predictor of performance on the PSAT 8/9 and PSAT 10, specifically prediction of status in relation to the College Board CCR Benchmarks. The current links are based on a small empirical validation study conducted outside of NM, which is one indicator of college and career readiness. Linking studies conducted specifically for NM students will provide the links necessary for monitoring student progress toward college and career readiness.</p>
Individual Students	
<p style="text-align: center;">Master Claim</p> <p>Performance on the NM-MSSA indicates a student’s progress toward college and career readiness.</p>	<p>College and career readiness requires that students can, without significant scaffolding, comprehend and evaluate grade-level problems in mathematics as set forth in the <i>New Mexico Common Core State Standards</i> for mathematics with connections to the Standards for Mathematical Practices.</p>



<p>Interpretations Using Proficiency Level Labels and Proficiency Level Descriptors (PLDs)</p> <p>Student scores coincide with one of four levels: Advanced, Proficient, Nearing Proficiency, and Novice.¹</p> <p>The PLD for each proficiency level describes what students are expected to know and be able to do in relation to the <i>New Mexico Common Core State Standards</i> for mathematics in grades 3–8. New Mexico students are expected to perform at the Proficient level to demonstrate mastery of the knowledge and skills needed to indicate college and career readiness.</p>	<p>A student’s proficiency level indicates how the student performed in relation to the knowledge and skills assessed in mathematics at that grade level.</p> <p>Proficiency level descriptors indicate the knowledge and skills that students are expected to be able to demonstrate at a level.</p>
<p>Interpretations Using Proficiency Level Descriptors</p> <p>A student’s proficiency level indicates that the student can be expected to demonstrate the knowledge and skills described at that level and in the levels below.</p>	<p>The student’s proficiency level also indicates that the student has mastered the knowledge and skills of the preceding proficiency levels.</p>
<p>Interpretations Using Scale Scores</p> <p>Scale scores provide a measure of student performance regardless of which form of the NM-MSSA is administered.</p>	<p>Scale scores indicate the student’s performance, regardless of which form of the NM-MSSA is taken.</p> <p>The proposed scale score reporting scale is under discussion.</p>
<p>Uses of Scale Scores</p> <p>Scale scores can be used to compare an individual student’s performance to the performance of other students in the school, district, and state.</p>	<p>Scale scores also indicate a student’s performance in relation to the performance of other students.</p> <p>A student’s scale score should be interpreted as the range of possible scores within the error band around that score, not only as a single number. (Other terms for “error band” include “margin of error” and “confidence interval.”)</p> <p>Differences between scale scores (e.g., for two students or a student’s score and a proficiency level cut score) that are within the margin of error should be interpreted as “statistical ties” (i.e., not reliably different).</p>
<p>Interpretation of Operations & Algebraic Thinking Grades 3–5 Indicators</p> <p>Student performance in the Operations & Algebraic Thinking indicators is reported in three levels:</p>	<p>Student performance on each of the Mathematics claims indicators is based on the items that target each of the claim areas. These items also explicitly target a depth of knowledge (DOK) level</p>



<p>Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient.²</p> <p>Performance on Operations & Algebraic Thinking items reflects application of mathematical operations; using algebraic representations to solve problems involving whole numbers; and identifying, explaining, and extending arithmetic patterns.</p>	<p>that categorizes the cognitive demand of the item content.</p>
<p>Interpretation of Number & Operations in Base Ten and Number & Operations – Fractions Grades 3–5 Indicators</p> <p>Student performance on Number & Operations in Base Ten and Number & Operations – Fractions indicators is reported in three levels: Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient.²</p> <p>Performance on Number & Operations in Base Ten items reflect understanding and using whole number place values to represent and interpret numbers, understanding the concept of fractions, representing fractions and decimal fractions, and comparing the sizes of whole numbers and fractions.</p>	<p>Student performance on each of the Mathematics claims indicators is based on the items that target each of the claim areas. These items also explicitly target a depth of knowledge (DOK) level that categorizes the cognitive demand of the item content.</p>
<p>Interpretation of Measurement & Data and Geometry Grades 3–5 Indicators</p> <p>Student performance on Measurement & Data and Geometry indicators is reported in three levels: Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient.²</p> <p>Performance on Measurement & Data and Geometry items reflects understanding measurement principles and applying them to solve problems, representing and analyzing data in simple graphs, and understanding geometric principles and using them to describe objects and solve problems.</p>	<p>Student performance on each of the Mathematics claims indicators is based on the items that target each of the claim areas. These items also explicitly target a depth of knowledge (DOK) level that categorizes the cognitive demand of the item content.</p>
<p>Interpretation of Ratios & Proportional Relationships Grades 6 and 7 Indicators</p> <p>Student performance on Ratios & Proportional Relationships indicators is reported in three levels: Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient.²</p> <p>Performance on Ratios & Proportional Relationships items reflects understanding, representing, and</p>	<p>Student performance on each of the Mathematics claims indicators is based on the items that target each of the claim areas. These items also explicitly target a depth of knowledge (DOK) level that categorizes the cognitive demand of the item content.</p>



<p>interpreting ratios and proportional relationships between variables to solve problems.</p>	
<p>Interpretation of Functions Grade 8 Indicators</p> <p>Student performance on Functions indicators is reported in three levels: Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient.²</p> <p>Performance on Functions items reflects understanding the concept of functions and representing linear functions in equations, tables, and graphs; comparing properties of two functions; and interpreting linear and nonlinear functions presented in a variety of forms.</p>	<p>Student performance on each of the Mathematics claims indicators is based on the items that target each of the claim areas. These items also explicitly target a depth of knowledge (DOK) level that categorizes the cognitive demand of the item content.</p>
<p>Interpretation of The Number System and Expressions & Equations Grades 6–8 Indicators</p> <p>Student performance on The Number System and Expressions & Equations indicators is reported in three levels: Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient.²</p> <p>Performance on The Number System items in grades 6 and 7 reflects understanding, representing, and computing with rational numbers.</p> <p>Performance on The Number System items in grade 8 reflects understanding and comparing rational and irrational numbers.</p> <p>Performance on Expressions & Equations items reflects using expressions, equations, and inequalities to represent and solve mathematical and real-world problems.</p>	<p>Student performance on each of the Mathematics claims indicators is based on the items that target each of the claim areas. These items also explicitly target a depth of knowledge (DOK) level that categorizes the cognitive demand of the item content.</p>
<p>Interpretation of Geometry and Statistics & Probability Grades 6–8 Indicators</p> <p>Student performance on Geometry and Statistics & Probability indicators is reported in three levels: Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient.²</p> <p>Performance on Geometry and Statistics & Probability items reflects understanding and applying geometric properties related to area, surface area, volume, and angles to solve real-world and mathematical problems; representing and analyzing data in a variety of plots and graphs; and summarizing and describing distributions using multiple measures.</p>	<p>Student performance on each of the Mathematics claims indicators is based on the items that target each of the claim areas. These items also explicitly target a depth of knowledge (DOK) level that categorizes the cognitive demand of the item content.</p>



<p>Interpretation of Problem Solving, Reasoning, and Argument Indicators</p> <p>Student performance on Problem Solving, Reasoning, and Argument indicators is reported in three levels: Met/Exceeded Proficient, Nearing Proficient, Did Not Meet Proficient.²</p> <p>Performance on Problem Solving, Reasoning, and Argument items reflects applying grade-level appropriate mathematical concepts and procedures and quantitative and logical reasoning to solve standard and nonstandard real-world and mathematical problems; and constructing viable arguments and critiquing the reasoning of others.</p>	<p>Student performance on each of the Mathematics claims indicators is based on the items that target each of the claim areas. These items also explicitly target a depth of knowledge (DOK) level that categorizes the cognitive demand of the item content.</p>
<p>Interpretation of Modeling, Patterns, and Structure Indicators</p> <p>Student performance on Modeling, Patterns, and Structure indicators is reported in three levels: Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient.²</p> <p>Performance on Modeling, Patterns, and Structure items reflects using grade-level appropriate quantitative reasoning to interpret mathematical representations, representing real-world mathematical situations using mathematical models, using mathematical models to solve real-world and mathematical problems, and looking for and making use of structure and repeated reasoning.</p>	<p>Student performance on each of the Mathematics claims indicators is based on the items that target each of the claim areas. These items also explicitly target a depth of knowledge (DOK) level that categorizes the cognitive demand of the item content.</p>
<p>Item Level Reporting for Individual Students</p> <p>Individual student performance on individual test items may suggest potential areas of strength and learning needs.</p>	<p>Caveat: Students may perform differently on items from other test forms that target the same subset of mathematics standards.</p>
<p>Groups of Students</p>	
<p>SIU statements for groups of students are applicable to aggregate reporting of school, district, and state performance and student subgroups (e.g., English learners, students with disabilities, racial/ethnic subgroups) within those levels of aggregation.</p>	
<p>Group Mean Scale Scores</p> <p>Group mean scale scores can be compared to other schools and districts, to the state, and for all students</p>	<p>Mean (i.e., average) scale scores enable comparison of performance among schools, districts, and other groupings of students.</p>



<p>and student subgroups (e.g., gender, English learners, students with disabilities).</p>	<p>Mean scale scores and percentages of students in a proficiency level for small groups (e.g., fewer than 25 students) are unstable and should be interpreted with caution because of concerns about reliability and stability.</p>
<p>Percentages of Students in Proficiency Levels Percentages of students in the four proficiency levels can be compared to other schools, districts, and the state, and for all students and student subgroups.</p>	<p>These are the percentages of students in each mathematics proficiency level. The PLD for each mathematics proficiency level indicates the degree of mastery of the knowledge and skills needed to indicate college and career readiness in relation to the <i>New Mexico Common Core State Standards</i>. The percentages of students in each level indicate the percentage of students who need to reach the next proficiency level. Means and percentages of students in a proficiency level for small groups should be interpreted with caution because of concerns about reliability and stability.</p>
<p>Item Level Reporting for Student Groups Student group performance (e.g., boys, girls, English learners) on individual test items may suggest potential areas of strength and learning needs.</p>	<p>Caveat: Students may perform differently on items from other test forms that target the same subset of mathematics standards.</p>

Unintended Score Interpretations and Uses

Until the NM-MSSA is in operational use, we can only speculate on what unintended interpretations and uses of NM-MSSA scores and other information may arise. Where unintended interpretations and uses may occur, it is the responsibility of that user to provide supporting evidence, and not the responsibility of PED (as specified in the *Standards for Educational and Psychological Testing*, 2014). The main concern for misinterpreting or misusing NM-MSSA scores is the potential negative consequences for individual students, subgroups of students, and schools, districts, and the state. If unintended interpretations and uses with potential negative consequences arise, PED will take steps to ameliorate the misinterpretations, misuses, and negative consequences. Some common misinterpretations and misuses that can arise include the following.

Interpreting Test Scores as 100% Accurate Indicators of Test Performance

All measurements in the real world, including test scores, are estimates. Test scores—for example, scale scores and proficiency level classifications—are estimates accompanied by a standard error. Standard errors are often referred to as the “margin of error” (e.g., in political polling). Interpreting and using NM-



MSSA scores correctly requires considering the width of the margin of error around a score. For example, students with a scale score 2 points below the cut score for the Proficient level could, hypothetically, have scored above the Proficient cut score on a different day because the NM-MSSA scale score standard errors are expected to be 2–3 points. Interpretations of NM-MSSA scores should account for the margin of error around each score estimate.

Drawing Conclusions and Making Decisions Based Solely on NM-MSSA Scores

There is wide agreement that conclusions and decisions based on a single piece of evidence can be risky. The risk is that the single piece of evidence can lead to less than optimal decisions, such as students failing to receive additional instruction based solely on their NM-MSSA score or teacher teams not being eligible for additional professional learning based solely on their students' NM-MSSA scores. Interpretations and uses of NM-MSSA scores should be supplemented with additional information.

Overinterpreting Subdomain Indicators and Item Level Performance Information

Subdomain indicators (e.g., Operations and Algebraic Thinking, Measurement and Data) are based on fewer items than are NM-MSSA total test scores. As a result, they are less stable estimates of student achievement and learning needs in that subdomain. In addition, because the performance indicators for the mathematics subdomains are highly correlated, differences in those performance indicators may be smaller than the proficiency level labels may suggest. Interpretations and uses of indicator scores should be supplemented with additional information.

Misinterpreting Current Performance as the Most Likely Predictor for Future Performance

A goal of education is to improve students' current achievement—that is, to bend their performance trajectory upward. It would be unwise to assume that students who currently are performing at the Proficient and Advanced levels will continue at these levels without sustained effort. Similarly, it would be unwise—and unfair—to assume that students who currently are performing at the Novice and Nearing Proficiency levels will perform at these levels in the future. In fact, our duty as educators is to help these students learn more and achieve higher.

Misinterpretations about students' current proficiency levels and future performance is not really a misinterpretation of NM-MSSA scores. It is a logical error in concluding that current performance determines future performance.

Overinterpreting NM-MSSA Scores as Indicators of College and Career Readiness

The New Mexico *Common Core State Standards* are designed to prepare students to be able to benefit from college study and postsecondary training. The claim that performance on NM-MSSA indicates readiness for college and career is supported only by the evidence contained in the mathematics content standards. NM-MSSA scores also can be interpreted as predictors of future performance in college and career training. However, until empirical prediction studies are completed, this



interpretation of NM-MSSA performance should be made with caution and with attention to the strong, but limited, evidence in the content standards.

¹ NM-MSSA Policy Proficiency Level Descriptors

Advanced. Students demonstrate evidence of **thorough** understanding and use of college and career readiness knowledge, skills, and abilities.

Proficient. Students demonstrate evidence of **satisfactory** understanding and use of college and career readiness knowledge, skills, and abilities.

Nearing Proficiency. Students demonstrate evidence of **partial** understanding and use of college and career readiness knowledge, skills, and abilities.

Novice. Students demonstrate evidence of **emerging** understanding and use of college and career readiness knowledge, skills, and abilities.

² NM-MSSA subscore indicators are reported as Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient. These subscore indicators are calculated using the IRT item parameters only for the items that are tagged for each subscore indicator. They are based on the cut score for Nearing Proficient, means, and conditional standard errors.