



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

Reading in English and Spanish

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, subdomain 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: Reading in English and Spanish

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

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

NM-MSSA score reports include scale scores for ELA, Reading, and Writing & Language. The ELA scale score includes performance on the Reading, Writing & Language, and Writing sections of the test. The Reading scale score is based only on reading items. The Writing & Language scale score includes both Writing & Language selected response items and the Writing task. Score reports for NM-MSSA Writing include only rubric scores (i.e., no scale scores).

**Intended Interpretations and Uses for Individual Students and Groups of Students: NM-MSSA
Reading in English and Spanish**

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</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>



Score Interpretation/Use Statement	Explanation/Annotation
<p>they have mastered <i>the New Mexico Common Core State Standards</i> for English language arts. 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>The reading standards require understanding key ideas and details, analyzing elements of craft and structure, and integrating knowledge and ideas in informational and literary texts. See details at https://webnew.ped.state.nm.us/bureaus/assessment-3/nm-mssa/.</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> <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) skills that require students to understand key ideas and details, analyze elements of craft and structure, and integrate knowledge and ideas in informational and literary texts, which is the progression for the next level of the reading curriculum. 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 complex texts across a range of types and disciplines as set forth in the <i>New Mexico Common Core State Standards</i> for English language arts.</p>
<p style="text-align: center;">Interpretations Using Proficiency Level Labels and Proficiency Level Descriptors</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 reading 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 reading 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 style="text-align: center;">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 style="text-align: center;">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 style="text-align: center;">Uses of Scale Scores</p> <p>Scale scores can be used to compare an individual student’s performance to the proficiency levels and 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</p>



	<p>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 Literary Text Indicators</p> <p>Student performance on Literary Text indicators is reported in three levels: Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient.²</p> <p>Performance on Literary Text items involves reading, comprehending, analyzing, and interpreting grade-level appropriate literary text; themes, text structures, points of view, key ideas and details; new and unfamiliar words; and comparing and contrasting two literary texts.</p>	<p>Student performance on each of the four Reading claims indicators (i.e., Literary Text, Informational Text, Comprehension, and Analysis and Interpretation) 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 Informational Text Indicators</p> <p>Student performance on Informational Text indicators is reported in three levels: Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient.²</p> <p>Performance on Informational Text items requires reading, comprehending, analyzing, and interpreting main ideas and details and new and unfamiliar words; and comparing and contrasting information, arguments, and evidence in two texts; all in grade-level appropriate informational text.</p>	<p>Student performance on each of the four Reading claims indicators (i.e., Literary Text, Informational Text, Comprehension, and Analysis and Interpretation) 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 Comprehension Indicators</p> <p>Student performance on Comprehension indicators is reported in three levels: Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient.²</p> <p>Performance on Comprehension items requires applying reading skills and strategies to comprehend central ideas and themes, identify supporting details, and determine meaning of words and phrases in grade-level appropriate literary and informational text.</p>	<p>Student performance on each of the four Reading claims indicators (i.e., Literary Text, Informational Text, Comprehension, and Analysis and Interpretation) 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 Analysis and Interpretation Indicators</p> <p>Student performance on Analysis and Interpretation indicators is reported in three levels: Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient.²</p> <p>Performance on Analysis and Interpretation items requires applying reading skills and strategies to grade-</p>	<p>Student performance on each of the four Reading claims indicators (i.e., Literary Text, Informational Text, Comprehension, and Analysis and Interpretation) 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</p>



<p>level appropriate literary and informational text to analyze ideas, events, and characters and examine relationships among text elements; interpret themes, purposes, claims, and evidence; determine and evaluate points of view; analyze the impact of literal and figurative language; analyze word choice; compare and contrast information and author methods in two texts; and make inferences and draw conclusions using evidence from the texts to support analyses and interpretations.</p>	<p>the cognitive demand of the item content.</p>
<p>Item Level Reporting for Individual Students 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 Reading 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., gender, English learners, students with disabilities, racial/ethnic subgroups) within those levels of aggregation.</p>	
<p>Group Mean Scale Scores Group mean scale scores can be compared to other schools, districts, and the state, and for all students and student subgroups (e.g., gender, English learners, students with disabilities, racial/ethnic subgroups).</p>	<p>Mean (i.e., average) scale scores enable comparison of performance among schools, districts, and other groupings of students. 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.</p>
<p>Percentages of Students in Proficiency Levels Percentages of students in proficiency levels 1–4 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 reading proficiency level. The PLD for each reading 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 each 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</p>



	be interpreted with caution because of concerns about reliability and stability.
<p>Item Level Reporting for Student Groups</p> <p>Student group performance (e.g., boys, girls, English learners) on individual test items or groups of 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 Reading standards.</p>

Unintended Score Interpretations and Uses

Until the NM-MSSA is in operational use, we only can 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., Literary Text, Comprehension) 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 reading 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.

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 English language arts 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.

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.

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

¹ NM-ASR 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 subdomain indicators are reported as Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient. These subdomain indicators are calculated by comparing a student's subdomain performance to the subdomain performance distribution of students who are just barely Nearing



Proficient on the total test, and using the standard deviation of that distribution to determine the Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient indicators.