STATEMENTS OF SCORE INTERPRETATIONS AND USES (SIUs) FOR THE NEW MEXICO ASSESSMENT OF SCIENCE READINESS (NM-ASR)

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 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-ASR, the phrase refers broadly to **test scores** (i.e., total test scale scores, subdomain scores), **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-ASR

Using this broad interpretation of the phrase, the intended score interpretations and uses for NM-ASR 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.

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. Score reports for NM-MSSA Writing include only rubric scores (i.e., no scale scores).

Intended Score Interpretations and Uses for Individual Students and Groups of Students

Score Interpretation/Use Statement	Explanation/Annotation
NM-ASR Program Purpose Statements	
Program Purpose Statement, Grade 11 NM-ASR The grade 11 NM-ASR is 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 the New Mexico *STEM Ready! Science Standards*, which require integration of Science and Engineering Practices, Disciplinary Core Ideas, and Crosscutting Concepts to explain phenomena and solve problems. Results	NM-ASR scores should be interpreted in relation to the New Mexico *STEM Ready! Science Standards* that are targeted by the assessment. *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

Score Interpretation/Use Statement	Explanation/Annotation
are presented using scale scores and proficiency levels. Proficient performance in grade 11 indicates both mastery of currently assessed grade level and preceding grades' expectations and progress toward college and career readiness.	higher education, without the need for remediation. ***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. ***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/bureaus/college-career-readiness/ Evidence to support this NM-ASR college and career readiness claim is in the New Mexico *STEM Ready! Science Standards*, which are based on the Next Generation Science Standards (NGSS). The NGSS "constructed each performance expectation by linking concepts and practices that build coherently over time throughout K–12, thereby helping to ensure that students who meet the NGSS will be prepared to succeed in science courses in both 2- and 4-year institutions" (see NGSS *Appendix C – College and Career Readiness* at https://www.nextgenscience.org/sites/default/files/resource/files/NGSS%20Appendix%20C%20Final%20072613.pdf).
Program Purpose Statement, Grades 5 and 8 NM-ASR Performance on the grade 8 NM-ASR indicates student mastery of grade levels 3–5 and 6–8 expectations for integration of Science and Engineering Practices, Disciplinary Core Ideas, and Crosscutting Concepts as presented in the standards, which is the progression for the next level of science curriculum, and is a predictor of being on track for college and career readiness.	The four explanations at grade 11 (above) apply in grades 5 and 8. In addition, performance on the grade 8 NM-ASR can be interpreted as a potential predictor of performance on the grade 11 NM-ASR (pending empirical validation), which is one indicator of college and career readiness. Performance on the grade 5 NM-ASR can be interpreted as a potential predictor of performance on the grade 8 NM-ASR (pending

Score Interpretation/Use Statement	Explanation/Annotation
Proficient performance in grades 5 and 8 indicates both mastery of currently assessed grade level and preceding grades' expectations and progress toward college and career readiness.	empirical validation), which is a predictor of college and career readiness at grade 11.
Individual Students	
Master Claim Performance on the NM-ASR indicates a student's progress toward college and career readiness.	College and career readiness requires that students can make sense of phenomena and solve real-world problems by applying and interconnecting scientific knowledge and skills as set forth in the New Mexico *STEM Ready! Science Standards*.
Interpretations Using Proficiency Level Labels and Proficiency Level Descriptors (PLDs) Student scores coincide with one of four levels: Novice, Nearing Proficiency, Proficient, and Advanced.[1] The PLD for each proficiency level describes what students can be expected to know and be able to do in relation to the New Mexico *STEM Ready! Science Standards* in grades 5, 8, and 11. New Mexico students are expected to perform at the Proficient level to demonstrate sufficient mastery of knowledge and skills needed to indicate college and career readiness.	A student's proficiency level indicates how the student performed in relation to the knowledge and skills assessed in science at that grade level. Proficiency level descriptors indicate the knowledge and skills that students are expected to be able to demonstrate at a level.
Interpretations Using Proficiency Level Descriptors A student's proficiency level indicates that the student can demonstrate the knowledge and skills described at that level and in the levels below.	The student's proficiency level also indicates that the student has mastered the knowledge and skills of the preceding proficiency levels.
Interpretations Using Scale Scores Scale scores provide a measure of student performance regardless of which form of the NM-ASR is administered.	Scale scores indicate the student's performance, regardless of which form of the NM-ASR is taken. The **proposed** scale score reporting scale is under discussion.

Score Interpretation/Use Statement	Explanation/Annotation
Uses of Scale Scores Scale scores can be used to compare an individual student's performance to the performance of other students in the school, district, and state.	Scale scores also indicate a student's performance in relation to the performance of other students. 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.") 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).
Interpretations of Practices and Crosscutting Concepts in Physical Sciences Student performance on this science subdomain is reported in three levels: Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient. [2]	Student performance in this science subdomain is based on items that target Disciplinary Core Ideas in Physical Sciences plus Science and Engineering Practices and/or Crosscutting Concepts. Because indicators for Physical, Life, and Earth and Space Sciences are likely to be highly correlated and will have non-trivial standard errors, proficiency levels for most students are likely to be identical in all three science subdomains.
Interpretations of Practices and Crosscutting Concepts in Life Sciences Student performance on this science subdomain is reported in three levels: Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient. [2]	Student performance in this science subdomain is based on items that target Disciplinary Core Ideas in Life Sciences plus Science and Engineering Practices and/or Crosscutting Concepts. Because indicators for Physical, Life, and Earth and Space Sciences are likely to be highly correlated and will have non-trivial standard errors, proficiency levels for most students are likely to be identical in all three science subdomains.
Interpretations of Practices and Crosscutting Concepts in Earth and Space Sciences	Student performance in this science subdomain is based on items that target Disciplinary Core Ideas in Earth and Space

Score Interpretation/Use Statement	Explanation/Annotation
Student performance on this science subdomain is reported in three levels: Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient.[2]	Sciences plus Science and Engineering Practices and/or Crosscutting Concepts. Because indicators for Physical, Life, and Earth and Space Sciences are likely to be highly correlated and will have non-trivial standard errors, proficiency levels for most students are likely to be identical in all three science subdomains.
Item Level Reporting for Individual Students Individual student performance on individual test items may suggest potential areas of strength and learning needs.	Caveat: Students may perform differently on items from other test forms that target the same subset of Science standards.

Groups of Students	
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.	
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).	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 because of concerns about reliability and stability.
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.	These are the percentages of students in each science proficiency level. The PLD for each science proficiency level indicates the degree of mastery of the knowledge and skills needed to indicate college and career readiness in relation to the *New Mexico STEM Ready! Science Standards*. 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

Score Interpretation/Use Statement	Explanation/Annotation
	interpreted with caution because of concerns about reliability and stability.
Item Level Reporting for Student Groups 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—with the caution that a student group may perform differently on other items that target the same Disciplinary Core Ideas, Science and Engineering Practices, and Crosscutting Concepts.	Caveat: Students may perform differently on items from other test forms that target the same subset of science standards.

Unintended Score Interpretations and Uses

Until the NM-ASR is in operational use, we can only speculate on what unintended interpretations and uses of NM-ASR scores and other information may arise. Where unintended interpretations and uses may be in use, 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-ASR 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-ASR 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-ASR scale score standard errors are expected to be 2–3 points. Interpretations of NM-ASR scores should account for the margin of error around each score estimate.

Drawing Conclusions and Making Decisions Based Solely on NM-ASR 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-ASR score or teacher teams not being eligible for additional science professional learning based solely on their students' NM-ASR scores. Interpretations and uses of NM-ASR scores should be supplemented with additional information.

Overinterpreting Subdomain Indicators and Item Level Performance Information

Subdomain indicators (e.g., Interpretations of Practices and Crosscutting Concepts in Life Sciences) are based on fewer items than are NM-ASR 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 three science 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. We assume that students who currently are performing at the Proficient and Advanced levels will continue at these levels only with sustained effort and support. 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-ASR scores. It is a logical error in concluding that current performance determines future performance.

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

The New Mexico *STEM Ready! Science Standards* are designed to prepare students to be able to benefit from college study and postsecondary training. The claim that performance on NM-ASR indicates readiness for college and career is supported only by the evidence contained in the science content standards. NM-ASR 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-ASR performance should be made with caution and with attention to the strong, but limited, evidence in the content standards.

---

[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-ASR 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 by using the standard deviation of that distribution to determine the Met/Exceeded Proficient, Nearing Proficient, and Did Not Meet Proficient indicators.