

Unified Academic Achievement Standards for Illinois State Assessments Policy Definition Workshop

Summary of Recommendations and Draft Assessment Policy Definitions Revised July 25, 2024

Introduction

The Illinois State Board of Education (ISBE) convened a broad-based group of education leaders from across Illinois on June 11, 2024, to provide guidance for establishing unified academic achievement standards for Illinois state assessments. In particular, ISBE is interested in establishing clear and coherent academic performance standards across its state assessment programs that reflect appropriate expectations for student achievement.

Staff from ISBE and the National Center for the Improvement of Educational Assessment (The Center) facilitated the workshop in Springfield. Workshop participants were asked to help develop policy definitions for state assessments and to provide recommendations to inform the development of performance level descriptors (PLDs), which will occur in a subsequent phase.

The group was reconvened virtually on July 18, 2024, to review draft policy definition levels, labels, and wording; to discuss these aspects of the policy definitions; and to voice additional perspectives to inform a version of the policy definitions to be shared by ISBE for public comment.

The purpose of this document is to summarize the panel's recommendations concerning:

- **Number and Name:** *How many performance levels should be set for state assessments and what should they be called?*
- **Rigor:** *How will ISBE know that the rigor of performance expectations is appropriate, particularly with respect to external benchmarks?*
- **Policy Definition:** *How should ISBE frame policy definitions that establish the vision for student academic achievement in Illinois?*
- **Coherence:** *What should ISBE consider to establish an appropriate relationship among performance levels across grades?*

These policy definitions and the associated guidance report will be used to provide an orientation to educators writing PLDs, which are more detailed descriptions of what students know and are able to do in relation to the Illinois Learning Standards. The policies and guidance in this report also will serve as a framework for evaluating the PLDs, ensuring that throughout the unified academic performance standard setting process, stakeholders can ask questions about whether the work is clear, at the appropriate level of rigor, expressed in ways that reflect



the values of stakeholders, and that support appropriate and effective use of state standardized assessment data.

Standardized summative assessment data are best suited to detect patterns in the performance of groups of students. They provide districts and schools with the opportunity to see where these patterns differ by grade, subject, demographic, or programmatic characteristics; prompt inquiry into the factors contributing to the differences; and suggest other types of information necessary to evaluate instructional programming and practices. They also provide a common standard across the state, situating school performance in a broader context so that districts and the state can identify areas of strength and opportunity, and consider how to strategically use resources. Last, as part of a set of multiple measures, state standardized assessment data can be used to consider the academic opportunities that might benefit a student and further support their long-term success.

Number and Name

The panel discussed the merits of having three, four, or five performance levels. The minimum number of performance levels permitted under federal regulation is three, where there must be a level at, above, and below proficient. A vast majority of states have four performance levels. Illinois currently has five performance levels on the Illinois Assessment of Readiness (IAR), so this option was presented as well. Some participants favored three levels because that places the focus on proficiency, and additional levels are not vital to the assessment's primary interpretation and uses.

However, a majority of panel members supported establishing four performance levels for all state assessment programs. At least one level would describe advanced achievement and another would correspond with "proficiency" (however named); the other two levels would describe performance that approaches proficiency.

In framing this recommendation, the panel reasoned that four levels would support consistency across assessment programs, meet federal requirements, and not require substantial changes to state tests. In particular, participants opposed lengthening any tests to support additional levels.

There was not a consensus recommendation regarding the name for performance levels, but the panel provided some guidance to inform the decision and some terms were proposed. The guidance included:

- Avoid deficit language. The names and descriptions should focus on what performance is demonstrated rather than what is deficient. Some terms, such as basic, does not meet, and unsatisfactory, were explicitly rejected due to negative connotations.
- Focus on performance, not students. The labels describe observed test performance and are primarily used to provide information for groups (e.g., schools, districts) and not



characteristics of students. Therefore, avoid any terminology that fails to honor this distinction.

- Prioritize clarity and utility. Names and descriptors that provide clear information to educators and other constituents are preferred over terms that are ambiguous or associated with “jargon.”

Some terms that gained support in whole or part from the group at the June 11 meeting included the following:

- Beginning (or Novice), Developing, Proficient, and Advanced (or Distinguished)
- [Not Identified], Approaching, Proficient, Mastery
- Did Not Meet, Partially Meets, Meets, Exceeds (or Surpasses)
- No names, only numbers: Level 1, Level 2, Level 3, and Level 4

Rigor

In the context of assessment performance levels, rigor refers to the knowledge and skills expected of a student to be considered proficient. Participants at the initial meeting overwhelmingly agreed that the current definition of “proficiency” for students is too high of an academic standard. In fact, 18 of 21 panel members agreed or strongly agreed with the statement, “Illinois’ definition of proficiency for students signals too high of an academic standard.”

Next, participants reviewed historical performance data from the Illinois Assessment of Readiness and Illinois Science Assessment as well as performance data from the ACT, SAT, and the National Assessment of Education Progress (NAEP). Reflecting on these data and their own experiences and expertise, they were asked to consider what, if any, external benchmarks signal that rigor is at or near an appropriate level. Small and large group discussions yielded the following recommendations:

- The state’s definition of “proficiency” should not be substantially different from other state assessment programs. (Subsequent polling on July 18 revealed less support for this recommendation. Discussion suggested not all states were equally suitable for comparison purposes.)
- The state’s definition of “proficiency” should not be substantially different from meaningful benchmarks set on national tests, such as ACT/SAT threshold for readiness (or on-track to readiness).
- There should be a positive and predictive relationship between demonstrated accomplishments associated with the academic demands related to postsecondary success (e.g., successful completion of Advanced Placement/dual credit coursework, internships, industry certifications) and proficiency on the state test. In other words, if most students who complete these accomplishments are not proficient on the state test, the rigor established for “proficiency” is likely too high.



- While participants did not identify separate on-track measures in Grades 3-8, they agreed that the progression of rigor across grades should increase in relatively consistent increments.

Policy Definition

Policy definitions are general statements that describe the intended level of rigor and desired performance within each performance level range. After reviewing several examples of state policy definitions, the panel discussed the characteristics and features that were most appropriate and useful for Illinois. In addition to affirming many of the recommendations from the *Number and Name* discussion, additional guidance included:

- Focus on characteristics of student performance with respect to the state’s academic content standards.
- Signal academic readiness or progress toward a range of postsecondary outcomes, including college and career.
- Promote clarity and utility by referencing some actions or next steps, when appropriate. However, be careful not to be too prescriptive. Avoid “diagnostic” interpretations that are not supported by a summative test.

More broadly, the panel was careful to note that academic achievement as measured on the state assessment is only one part of a much broader set of competencies associated with postsecondary success. Assessment results must be used in combination with other sources of evidence when supporting claims about readiness.

There was no clear consensus at the end of the initial (June 11) meeting on at least two issues:

- Inclusion of support language. One group of panelists preferred to avoid support language in the policy definitions altogether, consistent with their view that the definitions describe performance on a single occasion and cannot be the sole basis for subsequent actions. However, others thought that support language was appropriate because it makes the policy definitions more clear and useful.
- Referencing readiness for the next course of study. Some panelists felt that a single testing occasion could not support claims about readiness for the next course of study. The draft definitions respond to this concern.

At the follow-up meeting on July 18, the group revisited the discussion about naming. A majority (not all) favored the use of numbers for the levels, deemphasizing labels. Several additional considerations and concerns were raised, including:

1. Consider if language should be added or revised to avoid potential misinterpretation of level numbers (1 through 4) as corresponding to grades or GPA.



2. Clarify that statements associated with support do NOT refer to classification for special education services or similar consequences and that decisions about the nature of such support should be in conjunction with multiple measures.
3. Acknowledge that postsecondary success refers to a range of competencies that go beyond academics.

Based on this guidance and feedback from state examples regarded as more promising, the following policy definitions are presented as a draft for further review.

Level Range	Assessment Performance within this range ...
Level 4	Demonstrates advanced knowledge and application of the assessed Illinois Learning Standards for this subject and grade. Performance in the <i>Level 4</i> range is strong evidence that students are prepared for the academic demands of the next course of study and are progressing toward the academic expectations of college and career.
Level 3	Demonstrates proficient knowledge and application of the assessed Illinois Learning Standards for this subject and grade. Performance in the <i>Level 3</i> range is evidence that students are prepared for the academic demands of the next course of study and are progressing toward the academic expectations of college and career.
Level 2	Demonstrates a developing knowledge and application of the assessed Illinois Learning Standards for this subject and grade. Performance in the <i>Level 2</i> range is evidence that students may need additional support (the nature and manner of which should be informed by multiple sources of information) to demonstrate success with the academic demands of the next course of study or to accelerate progress toward the academic expectations of college and career.
Level 1	Demonstrates a beginning knowledge and application of the assessed Illinois Learning Standards for this subject and grade. Performance in the <i>Level 1</i> range is strong evidence that students may need additional support (the nature and manner of which should be informed by multiple sources of information) to demonstrate success with the academic demands of the next course of study or to accelerate progress toward the academic expectations of college and career.

Note: When in print, a footnote will indicate the Illinois Learning Standards may be accessed at <https://www.isbe.net/Pages/Standards-Courses.aspx>.

Readiness and support language is included in this draft, but in the context of a clause about the strength of the evidence that the test performance provides for these inferences.

Coherence

Finally, the panel was asked to provide feedback to help establish an appropriate relationship among performance levels across grades.

While acknowledging that some variation may be expected due to the nature of the state's learning standards, most participants proposed that the rigor across grades should be established such that it is relatively consistent. In other words, even if the "right" patterns of rigor across Grades 3-8 are difficult to identify, a very irregular progression across grades signals a "wrong" solution.

There was not widespread agreement on the expected impact (i.e., percent in performance level) that might be observed under these conditions. Many participants reasoned that even if rigor is set relatively consistently, it is realistic to observe performance declines as students progress from elementary grades to middle and high school. This may occur because students are introduced to a growing number of standards; any cumulative gaps in preparation will have a greater impact over time. This gradual decline may be particularly true for mathematics compared to English language arts. Other participants contended that performance should be stable across grades if the rigor is set appropriately. Still others expressed optimism that moderate gains over time could be observed with improvements in curriculum, instruction, and assessment.

Conclusion

The working group did not reach -- nor was it asked to reach -- unanimous agreement on these policy definitions, although most aspects were supported by two-thirds or more of the participants. When asked to share final thoughts, one member reflected that the lack of unanimity was not unexpected and that "we should expect the public to struggle in similar ways." Indeed, it is unlikely that any policy definition would equally satisfy all stakeholders, as there are diverse views on the most appropriate role and value, if any, for state standardized testing. Several members offered comments to this effect in their final thoughts, emphasizing the need to situate state standardized assessment as just one of multiple measures necessary to meet the needs of individual students. Others expressed appreciation for the process and opportunity to reevaluate expectations for proficiency, hoping others would see the considerations the group gave to the impact these policy definitions would have on students, families, and teachers. However, as one member noted, while this might be an improvement over the current situation, much work remains. This summary report and the draft of assessment policy definitions serve to prompt comments and suggestions for improvement on the proposed assessment policy definitions, which will serve as the foundation for the next phase of the unified standard-setting process and the development of detailed content and range performance level descriptors.