

Oklahoma 2024 Standard Setting Evaluation Report

ACS Ventures, LLC

Susan Davis-Becker, Ph.D.
Chad Buckendahl, Ph.D.
Sarah Alahmadi, Ph.D.
Kyle Killworth, MS. Ed.

WestEd

Priya Kannan, Ph.D.
Sarah Quesen, Ph.D.

CONTENTS

Foreward and Glossary	3
Executive Summary	5
<i>Standard Setting Evidence Collection.....</i>	<i>6</i>
<i>Evaluation Findings</i>	<i>6</i>
<i>Recommendations</i>	<i>7</i>
Introduction	10
Key Recommendations	11
Evaluation Framework and Process	14
<i>Evaluation Framework</i>	<i>14</i>
<i>Sources of Information.....</i>	<i>15</i>
Evaluation Findings	17
<i>Design.....</i>	<i>17</i>
<i>Execution</i>	<i>18</i>
<i>Outcomes</i>	<i>18</i>
<i>Standard Setting</i>	<i>18</i>
<i>Comparison of Impact and Expectations</i>	<i>20</i>
<i>Summary of 2024 Standard Setting Evaluation</i>	<i>27</i>
Next Steps and Recommendations	28
References.....	30
Appendix A: Evaluation Questions, Information, and Findings by Phase.....	32
Appendix B: Oklahoma PLD Comparisons between 2017 and 2024	37
Appendix D: Proficiency Rates for Neighboring States	40

Foreward and Glossary

This report was developed for the Oklahoma Commission on Educational Quality and Accountability (CEQA) as a part of the Evaluation of the 2024 standard setting and related activities for the Oklahoma School Testing Program (OSTP). In this section we provide descriptions of the terms related to standard setting (general and technical) and how they are used in this report.

- **Academic Content Standards** – established at the state level, these standards outline the knowledge and skills that should be included in the instruction and assessment¹.
- **Academic Performance Standards / Cut Scores** – thresholds used to differentiate levels of academic performance on an assessment (e.g., a cut score of X for performance level A indicates that students who achieve a score of X or higher will be in performance level A).
- **Frame of Reference** – the information and resources provided to standard setting panelists to guide their judgments in a standard setting task. For example, this may include performance level descriptors or estimated pass rates with or without benchmarks (e.g., historical data, comparisons with other assessments).
- **Impact** – the effect of cut scores on student performance. In other words, the estimated percentage of students who would be classified in each performance level if a given set of cut scores were applied.
- **Judgmental Task** – the key activity within a standard setting method that defines what panelists (e.g., teachers) are asked to do (i.e., make judgments) to share their expertise and formulate recommendations.
- **Panelist** – individual who participates in a standard setting process by using their experience and expertise to make judgments about the expectations for student performance on the assessment.
- **Performance Levels** – categories of student performance that can be used to classify students in reporting assessment results. Oklahoma has four performance levels: Below Basic, Basic, Proficient, and Advanced. Other assessment programs (e.g., other states, NAEP) may define between three and five levels of student performance.
- **Performance Level Descriptors (PLDs)** – set of expectations for students at each performance level that describe what they know and are able to do. Relevant to standard setting, there are two types of PLDs:
 - **Policy PLDs** - Broad statements that define the intended rigor of each performance level across grades and subject areas. In Oklahoma, these are defined in statute (70 OK Stat § 1210.541 (2024)) as Below Basic, Basic, Proficient, and Advanced.
 - **Range PLDs** – Detailed statements that outline content-based expectations for students in a specific subject and grade, and for each performance level.
- **Proficiency Rate** - the percentage of students who are identified to be in the “proficient” performance level or higher (e.g., proficient + advanced).
- **Standard Error** (of Panelist Recommendations) – this value reflects the variability in panelists’ recommended cut scores. While some variability is expected due to differences in experience and areas of expertise, these values are reviewed for reasonableness to determine if panelists

¹ <https://oklahoma.gov/education/services/standards-learning/oklahoma-academic-standards.html>

were generally similar in their understanding of the standard setting task and how their resulting recommendations.

- **Standard Setting (Process)** – typically organized as a workshop, this is an event where expert panelists recommend cut scores through an iterative judgmental process.
- **Standard Setting Judgments** – See *Judgmental Task*
- **Standard Setting Results** (and Recommendations) – the outcome of the standard setting process including panelist judgments and the final recommended cut scores.
- **Technical Advisory Committee (TAC)** – a group with technical expertise in assessment and accountability, empaneled by the state, to provide oversight and advice on technical matters of assessment development and validation related to the state testing program.
- **Vertical Articulation** – the process of reviewing the impact of cut scores across grades within a subject and adjusting for coherence within the systems of cut scores.

Executive Summary

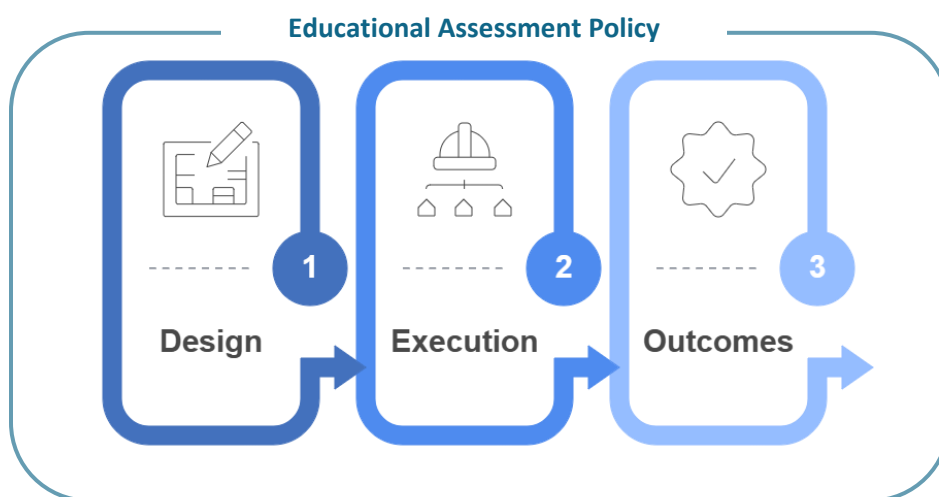
In 2025, the Oklahoma Office of Educational Quality and Accountability (OEQA) contracted with ACS Ventures (ACS) and WestEd to conduct an independent evaluation of the standard setting process used to establish cut scores for the Oklahoma School Testing Program (OSTP) in English Language Arts (ELA) and mathematics for grades 3–8 that was conducted in 2024. The goals of this evaluation were to:

- evaluate the design, execution, and outcomes of the 2024 standard setting and provide evidence to inform the CEQA’s decisions regarding the cut scores from this process (pursuant to 70 OK Stat § 1210.508 (2024)), and
- provide recommendations regarding future standard setting activities.

This evaluation was guided by professional testing standards (AERA et al., 2014) in educational assessment standard setting and spanned the policy, content, and empirical elements of the design of the standard setting, execution of the process, and the interpretation and use of the outcomes.

This evaluation was conducted following a model of standard setting and related activities shown in Figure 1. There are three phases of the process which are all nested within the larger educational assessment policy that specifies the purpose of the assessment and the intended interpretation of scores. The first phase is the design of the standard setting study in which standard setting experts work with policy and content experts to determine how the process will be conducted to ensure the outcomes reflect the purpose of the assessment and the intended interpretation of scores. The second phase of execution involves gathering input from content experts on the expectations for the intended examinees following the design laid out in the first phase. The third phase is focused on the outcomes of the execution phase, how these are evaluated, and put into practice in alignment with the intended interpretation of scores.

Figure 1: Phases of Standard Setting



Standard Setting Evidence Collection

Several key activities were identified within the design of this evaluation to accomplish the stated goals:

- **Review 2024 Standard Setting Documentation** – The evaluation team reviewed evidence that was publicly available about the 2024 standard setting. Additional materials were requested including the standard setting plan documentation, detailed results of the study, and feedback received from various stakeholder groups. However, this supplemental information was not made available to the evaluation team.
- **Stakeholder Interviews** – The evaluation team sought the opportunity to speak with individuals involved in the 2024 standard setting to learn more about specific decisions and their perceptions of the process including policy makers, technical advisors, standard setting vendor staff, and some of the expert panelists. The evaluation team was able to speak with one person from OSDE about the 2024 standard setting policy (Todd Loftin), members of the 2024 CEQA (Secretary Nellie Sanders, Dr. Ashlyn Fiegenger), and several individuals involved with the previous (2017) standard setting study² as SDE employees, members of the state Technical Advisory Committee (TAC), or in other advisory roles to the process. Other individuals did not respond to requests for participation or cited nondisclosure agreements for not being able to participate.
- **Comparison of Oklahoma Expectations to External Sources** – The evaluation team compared Oklahoma’s performance expectations qualitatively (e.g., Performance Level Descriptors [PLDs]) and resultant student proficiency rates to student performance regionally from neighboring states (i.e., Arkansas, Colorado, Kansas, Missouri, New Mexico, and Texas) and nationally with the National Assessment of Educational Progress (NAEP).

Evaluation Findings

In alignment with the goals of this evaluation and based on the information that was available when preparing this interim report, two key findings are presented here.

First, the 2024 standard setting processes did not follow the legislated requirements for CEQA to “establish panels to review and revise the performance level descriptors for each subject and grade level” and to “determine the cut scores for performance levels on all statewide assessments” (70 OK Stat § 1210.508 (2024)) which implies the need for the CEQA to be fully involved in the design, execution, and outcomes phases of the standard setting process. Rather, CEQA was only involved when the final results from the standard setting panel were presented for approval.

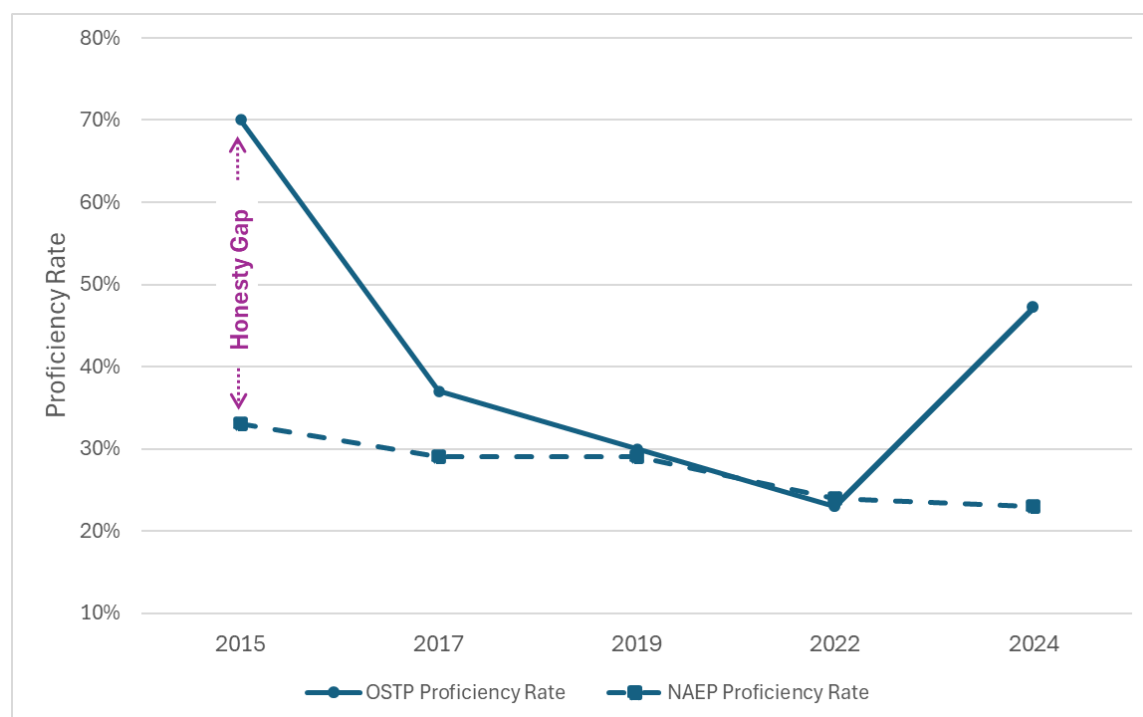
Second, the 2024 standard setting process seemed to largely rely on the range PLDs as the frame of reference for the standard setting judgment (i.e., no other clear policy influence). This is a notable difference from the previous standard setting (2017) in which there was a specific goal in Oklahoma to align the content standards and OSTP results with national standards such as the National Assessment of Educational Progress (NAEP) and college readiness measures (i.e., ACT, SAT; see Oklahoma State Department of Education, 2017). This 2017 policy goal followed Oklahoma adopting a term used by

² Although this evaluation was focused on the 2024 standard setting, a key finding was the notable contrast with the 2017 results. Therefore, the evaluation also included gathering information about that process to understand the policy perspective and how it was operationalized. However, many of these individuals asked to remain anonymous.

multiple states, “Honesty Gap,” (see Chu, 2025, Honesty Gap, n.d., Virginia Department of Education, 2022) to describe the difference between their proficiency rates and those reported by NAEP. The resulting performance standards produced proficiency rates that aligned with those reported by NAEP. In 2024, there did not appear to be any policy directive informing the standard setting process (including a plan to align with NAEP) and therefore, the Honesty Gap re-emerged.

Figure 2 shows the proficiency rate for Oklahoma students in grade 4 English language arts (ELA) as reported by NAEP and Oklahoma across five NAEP assessment years. In 2015, before the previous standard setting, the Honesty Gap was large but then notably reduced when the new cut scores were put into place in 2017. Smaller Honesty Gaps were observed in 2019 and 2022 but then the Honesty Gap widened in 2024 when the most recent cut scores were employed.

Figure 2. Change in Oklahoma ELA Honesty Gap Over NAEP Administrations



Recommendations

A synthesis of the findings yielded two recommendations in addition to some policy options for the CEQA:

First, we recommend that potential adjustments to the cut scores from 2024 be informed by policy goals, external data, and vertical articulation data. Because standard setting is an integration of policy, content, and data, it is important to ensure that the system of performance standards and cut scores is aligned with other aspects of the assessment system. Whether adjustments should or should not be made is ultimately a policy decision that considers the available evidence shared in this evaluation.

Based on the findings in this report, the CEQA is encouraged to consider the following policy options with respect to applying the results of the 2024 standard setting to the 2025 OSTP administration.

OPTION A: Adjust 2024 cut scores to apply to the 2025 administration – Because alignment with policy goals is an important consideration, if the CEQA determines that there was a lack of policy alignment with outcomes of the 2024 standard setting, they could adjust the cut scores before applying them to the 2025 OSTP administration. The policy-guided nature of any adjustment should be made in such a way that it still ensures alignment with the stated performance expectations (i.e., PLDs) and based on empirical evidence (e.g., student performance data, benchmarks). It is important to note that making an adjustment to cut scores that apply to the 2025 OSTP administration has downstream effects on score reporting and the state’s assessment vendor responsible for producing score reports as well as activities related to statewide accountability reporting.

OPTION B: Maintain the 2024 cut scores for the 2025 administration and implement any adjustments to cut scores for the 2026 administration – If there is not a clear policy goal against which to evaluate the cut scores or make adjustments for the current administration, the CEQA should reinforce its policy goals, communicate them to stakeholders, and use the results of this evaluation and the standard setting technical report to determine if adjustments to cut scores are necessary for the 2026 OSTP administration. This option provides an opportunity to coordinate on policy goals and would allow for adequate notice to the field regarding implementation of any changes.

OPTION C: Maintain 2024 cut scores with explanation of the policy rationale – As discussed in this report, our review of the available information about the 2024 standard setting suggests that it was conducted using an acceptable process. There are some areas of our evaluation where questions were raised that could not be answered with the available information. However, these questions do not necessarily represent concerns about the selection or execution of the standard setting methodology. Rather, our evaluation indicated that the strong policy perspective about alignment with national standards that influenced the 2017 cut scores did not seem to be a consideration in the 2024 standard setting process. This observation suggests a policy shift; however, this shift was not intended or recommended by the CEQA. Alternate policy interpretations may be reasonable but would benefit from clear communications about the frame of reference and rationale for each perspective and be determined by the CEQA as the policy body responsible for setting the cut scores.

Second, future standard setting processes must involve the CEQA in their legislatively authorized roles, as well as those with expertise in areas related to standard setting, to provide guidance regarding the design, execution, and outcomes. A key aspect of the validity argument supporting the use of assessment results is the alignment of components within the larger assessment system. With respect to standard setting, the cut scores should reflect the intended interpretation of assessment results as outlined in the performance level descriptors (PLDs). Given the need for these elements to be aligned, it is logical that the CEQA is legislatively responsible for both the PLDs (narrative descriptions of student performance expectations) and the cut scores (empirical representations of student performance expectations). To ensure this alignment, the CEQA should be involved throughout the standard setting processes (either setting new cut scores or validating existing cut scores) as follows.

- **PLD Development:** The CEQA should provide leadership in the development of the range PLDs from the policy PLDs (established through legislation). More specifically, the CEQA can identify relevant benchmarks such as alignment with national achievement indicators and references to how expectations can be operationalized across grades and subjects as well as identify those who should be involved in creating the range PLDs. The CEQA should collaborate with those with expertise in these activities to design and lead workshops following the policy guidance provided.
- **Standard Setting Design:** The CEQA should be directly involved in the design of the standard setting activities in collaboration with the standard setting experts/study leaders and the Oklahoma State Board of Education (SBE). More specifically, the CEQA should be involved in identifying the frame of reference and benchmarks that will help guide the panelists' judgments, creating a recruitment strategy for expert panelists, and developing the overall process for how feedback information will be used to inform panelist judgments.
- **Standard Setting Execution:** The CEQA should observe the execution of the standard setting to ensure the process is executed as outlined in the agreed-upon design, evaluate any deviations from the design, and be available to discuss any policy-related questions that arise with the standard setting leaders.
- **Standard Setting Outcomes:** The CEQA should continue their role of approving the cut scores but with additional knowledge and understanding of how their policy goals informed all aspects of standard setting, starting from the PLD development process, the design of the standard setting, the execution of the standard setting process, identification of relevant benchmarks against which the impact of the standard setting results can be reviewed, and the vertical articulation of the results to inform a coherent system of cut scores across grades and subject areas. The benchmark data that standard setting vendors often include in presentations to policymakers are the range of recommendations from panelists, historical student performance, current student performance, national indicators, and process evaluation from panelists that communicate their experiences with the process.

Introduction

The Oklahoma Office for Educational Quality and Accountability (OEQA) contracted with ACS Ventures, LLC (ACS) and WestEd to evaluate the 2024 standard setting and related activities for Oklahoma School Testing Program (OSTP). These activities were focused on establishing cut scores for the statewide assessments in grades 3-8 for English language arts (ELA) and mathematics. The purpose of this evaluation was to provide an independent review of the design, execution, and outcomes of the 2024 standard setting relative to Oklahoma's policy goals. In addition, this report includes recommendations to the Commission on Educational Quality and Accountability (CEQA) regarding standard setting policy, application of 2024 cut scores to the 2025 spring statewide assessment results, and future standard setting studies.

As an integration of policy, content, and data, standard setting is an important source of validation evidence for assessment programs that are used for summative purposes such as determining levels of student achievement, contributing to school and district accountability systems, professional development, informing curricular and instructional planning, and supporting student remediation. The process involves careful design, execution, and review of the outcomes to establish cut scores that should represent the operationalization of the statewide achievement standard policy for each grade and subject.

Oklahoma Statute §70-1210.508³ describes several characteristics of the statewide assessment system including such requirements as alignment with the state academic content standards and comparability among several states. For this evaluation, we focused on the characteristics associated with setting achievement standards. Interested readers are encouraged to review the full statute for the comprehensive expectations. This statute also identifies several policy stakeholders and the distributed responsibilities for aspects of the statewide assessment system that are shown here:

- **State Legislature** – Define requirements for assessment system as noted in the Statute.
- **State Board of Education** – Adopt assessments that meet requirements and retain independent experts with educational testing experience who do not develop aptitude or achievement tests for elementary or secondary students.
- **State Department of Education (SDE)** – Implement assessments and conduct annual evaluations of results.
- **Commission for Educational Quality Assurance (CEQA)** – Establish panels to review and revise the performance level descriptors for each subject and grade level and determine cut scores on statewide assessments and conduct ongoing reviews of comparability to other states.

As organizations independent of the standard setting studies conducted for the OSTP, ACS and WestEd collaboratively conducted the review of the 2024 standard setting, related activities, and comparability to external benchmarks. This review followed a set of guiding principles based on professional testing standards (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014) and best practices articulated in seminal references for the field (e.g., Hambleton & Pitoniak, 2006; Ferrara et al., in press). The principles that guided this evaluation were:

³ 70 OK Stat § 1210.508 (2024)

- Standard setting is part of the overall test development and validation process and should be conducted in alignment with the policy goals and design of the assessment.
- The policy goals guiding standard setting studies vary by state and assessment depending on the intended interpretation and use of scores. It is important to note that rendering judgment on the policy goals is not within the purview of this evaluation. Rather, this evaluation is to determine whether the design, execution, and outcomes of the study align with Oklahoma’s policy goals.
- There are not “correct” or “true” cut scores to be identified through a standard setting study. Rather, standard setting involves an integrated set of processes involving policy (see policy stakeholders noted above), content (subject matter experts), and data (empirical evidence to inform judgment).
- Operationally, a standard setting study for statewide educational assessments is often implemented using a systematic, judgmental process where subject matter experts (educators) make judgments for what they believe is reasonable in accordance with achievement standard policy, design of the assessment, their understanding of assessment content, and their understanding of expectations of student achievement (i.e., PLDs) for a given assessment.
- There are several methods for recommending cut scores that are acceptable practice for a given assessment (see Hambleton & Pitoniak, 2006). Therefore, this evaluation focused on how the method was operationalized (design and execution), how the results (outcomes) were reviewed, and final cut scores were selected.

In the next section of this report, we highlight our key recommendations for the CEQA in their review of the current cut scores. From there, the report provides details on the full scope of the evaluation including the 2024 standard setting and related activities and the comparison of the impact of this activity with results from several states and national benchmarks (NAEP). Next, the report provides a historical policy perspective along with information on the impact of the previous (2017) standard setting to understand the larger context. Finally, the fourth section outlines the next steps for the CEQA.

Key Recommendations

In this section, we highlight our key recommendations related to the current cut scores (set in 2024) for the OSTP grades 3-8 in ELA and mathematics.

First, we recommend that potential adjustments to the cut scores from 2024 be informed by policy goals, external data, and vertical articulation data. Because standard setting is an integration of policy, content, and data, it is important to ensure that the system of performance standards and cut scores is aligned with other aspects of the assessment system. Whether adjustments should or should not be made is ultimately a policy decision that considers the available evidence shared in this evaluation.

Based on the findings in this report, the CEQA is encouraged to consider the following policy options with respect to applying the results of the 2024 standard setting to the 2025 OSTP administration.

OPTION A: Adjust 2024 cut scores to apply to the 2025 administration – Because alignment with policy goals is an important consideration, if the CEQA determines that there was a lack of policy alignment with outcomes of the 2024 standard setting, they could adjust the cut scores before applying them to the 2025 OSTP administration. The policy-guided nature of any

adjustment should be made in such a way that it still ensures alignment with the stated performance expectations (i.e., PLDs) and based on empirical evidence (e.g., student performance data, benchmarks). It is important to note that making an adjustment to cut scores that apply to the 2025 OSTP administration has downstream effects on score reporting and the state's assessment vendor responsible for producing score reports as well as activities related to statewide accountability reporting.

OPTION B: Maintain the 2024 cut scores for the 2025 administration and implement any adjustments to cut scores for the 2026 administration – If there is not a clear policy goal against which to evaluate the cut scores or make adjustments for the current administration, the CEQA should reinforce its policy goals, communicate them to stakeholders, and use the results of this evaluation and the standard setting technical report to determine if adjustments to cut scores are necessary for the 2026 OSTP administration. This option provides an opportunity to coordinate on policy goals and would allow for adequate notice to the field regarding implementation of any changes.

OPTION C: Maintain 2024 cut scores with explanation of the policy rationale – As discussed in this report, our review of the available information about the 2024 standard setting suggests that it was conducted using an acceptable process. There are some areas of our evaluation where questions were raised that could not be answered with the available information. However, these questions do not necessarily represent concerns about the selection or execution of the standard setting methodology. Rather, our evaluation indicated that the strong policy perspective about alignment with national standards that influenced the 2017 cut scores did not seem to be a consideration in the 2024 standard setting process. This observation suggests a policy shift; however, this shift was not intended or recommended by the CEQA. Alternate policy interpretations may be reasonable but would benefit from clear communications about the frame of reference and rationale for each perspective and be determined by the CEQA as the policy body responsible for setting the cut scores.

Second, future standard setting processes must involve the CEQA in their legislatively authorized roles, as well as those with expertise in areas related to standard setting, to provide guidance regarding the design, execution, and outcomes. A key aspect of the validity argument supporting the use of assessment results is the alignment of components within the larger assessment system. With respect to standard setting, the cut scores should reflect the intended interpretation of assessment results as outlined in the performance level descriptors (PLDs). Given the need for these elements to be aligned, it is logical that the CEQA is legislatively responsible for both the PLDs (narrative descriptions of student performance expectations) and the cut scores (empirical representations of student performance expectations). To ensure this alignment, the CEQA should be involved throughout the standard setting processes (either setting new cut scores or validating existing cut scores) as follows.

- **PLD Development:** The CEQA should provide leadership in the development of the range PLDs from the policy PLDs (established through legislation). More specifically, the CEQA can identify relevant benchmarks such as alignment with national achievement indicators and references to how expectations can be operationalized across grades and subjects as well as identify those who should be involved in creating the range PLDs. The CEQA should collaborate with those

with expertise in these activities to design and lead workshops following the policy guidance provided.

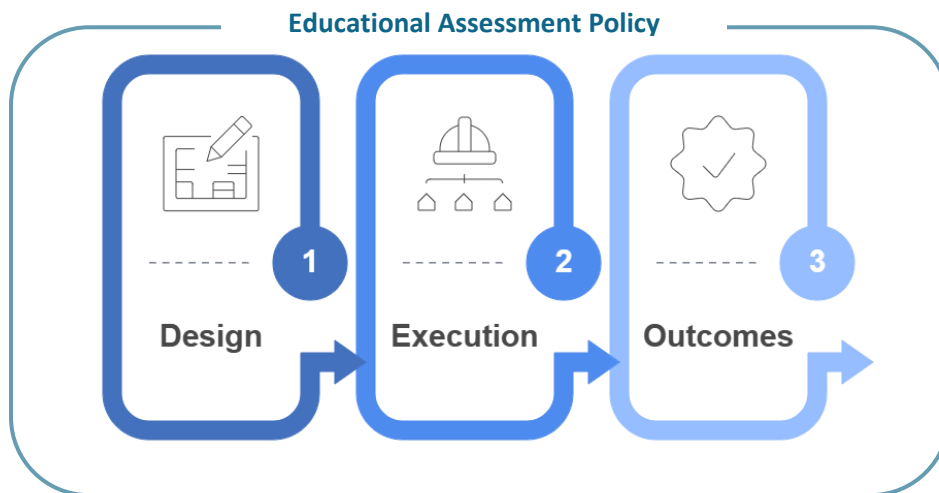
- **Standard Setting Design:** The CEQA should be directly involved in the design of the standard setting activities in collaboration with the standard setting experts/study leaders and the Oklahoma State Board of Education (SBE). More specifically, the CEQA should be involved in identifying the frame of reference and benchmarks that will help guide the panelists' judgments, creating a recruitment strategy for expert panelists, and developing the overall process for how feedback information will be used to inform panelist judgments.
- **Standard Setting Execution:** The CEQA should observe the execution of the standard setting to ensure the process is executed as outlined in the agreed-upon design, evaluate any deviations from the design, and be available to discuss any policy-related questions that arise with the standard setting leaders.
- **Standard Setting Outcomes:** The CEQA should continue their role of approving the cut scores but with additional knowledge and understanding of how their policy goals informed all aspects of standard setting, starting from the PLD development process, the design of the standard setting, the execution of the standard setting process, identification of relevant benchmarks against which the impact of the standard setting results can be reviewed, and the vertical articulation of the results to inform a coherent system of cut scores across grades and subject areas. The benchmark data that standard setting vendors often include in presentations to policymakers are the range of recommendations from panelists, historical student performance, current student performance, national indicators, and process evaluation from panelists that communicate their experiences with the process.

Evaluation Framework and Process

Evaluation Framework

This evaluation was conducted following a model of standard setting and related activities as shown in Figure 1. The first phase is the design of the standard setting study, the second phase is the execution of the standard setting process, and the third is the use of the standard setting outcomes for establishing the final cut scores. These phases are all nested within the larger educational assessment policy.

Figure 1: Phases of Standard Setting



This section provides an overview of each phase of the standard setting process and the critical areas of evaluation within. Appendix A details the evaluation questions related to each phase.

Design phase. In this phase, standard setting experts work with policymakers to determine how the standard setting will be conducted. Standard setting is a process where policy goals are operationalized so that subject matter experts can make systematic, iterative judgments about the expectations for students for a given grade and subject. The judgments include scaffolded information of content, policy expectations (e.g., purpose of the assessment, meaning of the cut scores, PLDs), and student data. Typically, evaluation in this area is conducted before the process continues into the execution phase but it is included here to the extent that the design elements could be parsed from the execution and outcomes phase.

The review of the design phase is largely focused on ensuring the execution phase has the capacity to be successful in allowing expert panelists to make judgments in accordance with the standard setting method and the policy goals for establishing cut scores. Specifically, the evaluation team sought to review the policy goals specified for the standard setting and interpretation of scores, how the policy goals were operationalized as a frame of reference for the panelists, the strategy for recruiting and selecting a representative group of qualified panelists, the selected standard setting method, what empirical information was included in the standard setting process, what benchmarks were provided to panelists to evaluate their judgments and recommendations, and whether panelists would be able to leverage their expertise to provide meaningful recommendations.

Execution phase. In this phase, standard setting experts conduct the standard setting process in accordance with the design outlined in the first phase. This typically includes an opportunity for panelists to learn about the purpose of the standard setting, the policy goals / frame of reference they will apply in the standard setting, the process and method they will follow, how they will make their judgments, and what information they will be provided to refine their judgments.

The review of the execution phase focuses on how well the process followed the design, how any deviations were handled and documented, and the panelists' perceptions of the process. Specifically, the evaluation sought to review the details of the workshop, the success of the training activities, and the panelists' understanding of their task and the information associated with the task.

Outcomes phase. In this phase, standard setting experts and policymakers review the results to determine the final cut scores. This may include one or more stages where groups with expertise in these areas provide input on the results. The final decision is typically the responsibility of a policy body who is provided with the information they need to thoroughly understand the design and execution of the study and evaluate the outcomes against the policy goals and relevant benchmarks.

The review of the outcomes phase is focused on the evidence that the standard setting results reflect the intended frame of reference, following a successful standard setting process. Specifically, the evaluation team sought to review evidence of the panelists' shared understanding of the frame of reference, their perception of the results, the consistency in their final results, and the reasonableness of the results.

Sources of Information

At the outset of the project, the evaluation team outlined three major sources of information for the evaluation. Each is described here along with information on which of these elements the evaluation team was able to obtain and evaluate.

2024 Standard Setting Documentation. The goal with this first activity was to review all available documentation that described the inputs to, and the results of, the design, execution, and outcomes phases. Some of this information was available publicly including the 2024 standard setting report, CEQA meeting minutes where standard setting activities were discussed, and Oklahoma legislation related to the OSTP cut scores. Additional information requested, but not received, included:

- 2017 standard setting report
- 2024 policy about the statewide assessments applicable to the standard setting process
- 2024 PLD development process report
- Additional information on the 2024 standard setting
 - Standard setting plan documentation and feedback on standard setting plan from technical or policy stakeholders
 - Content-based benchmark report
 - Detailed results (panelist level, by round) from the 2024 standard setting as well as feedback presentations provided after each round.
 - Standard setting observer reports
 - Standard setting results presentation to the TAC as well as TAC feedback on the results
- Public announcement of statewide results

Stakeholder Interviews. Given that the standard setting report (and other documentation) provides summaries of some related activities, ACS sought the opportunity to speak with individuals involved in the 2024 standard setting to learn more about specific decisions and their perceptions of the process. Specifically, the evaluation team attempted to speak with the individuals from the Oklahoma State Department of Education (SDE), the Oklahoma technical advisory committee (TAC), the vendor who designed and ran the 2024 standard setting (Cognia), and some of the experts who participated in the standard setting as panelists. The evaluation team was able to speak with one person from SDE (Todd Loftin), members of the 2024 CEQA (Secretary Nellie Sanders, Dr. Ashlyn Fiegenger), and several individuals involved with the prior standard setting study (2017) as SDE employees, members of the state Technical Advisory Committee, or in other advisory roles to the process⁴.

However, some individuals who were invited did not respond to, or accept, requests (current SDE staff: Superintendent Ryan Walters, Cathy Boomer, Alyssa Tyra, Corrine Baesler, and Rebecca Logan) whereas others cited nondisclosure agreements that were signed to protect the confidential aspects of the standard setting activity and indicated they could participate if this was waived (TAC members, standard setting observers, vendor staff). The requested information would have informed the evaluators of the policymakers' interpretation of the policy goals and execution of the standard setting activities. Because interviews with these stakeholders were not possible and the evaluation was largely constrained to publicly available documentation about the study in the technical report, gaps in the report were then necessarily interpreted as gaps in the design (e.g., policy goals, alignment with previous studies, integration of external evidence, rationale for inclusion/exclusion of information, roles of each policy body) and execution of the studies that produced the results.

Comparison of Oklahoma Expectations to External Sources. As part of the project, the evaluation team compared Oklahoma's performance expectations and resultant student proficiency rates to student performance regionally from neighboring states (i.e., Arkansas, Colorado, Kansas, Missouri, New Mexico, and Texas) and nationally with the National Assessment of Educational Progress (NAEP). This part of the analysis connected with the current Oklahoma legislation requiring that the assessments provide compatibility with other states (70 OK Stat § 1210.508 (2024)) and policy guidance that specifies Oklahoma assessments should be compared to national assessments (Oklahoma Edge, 2017). Publicly available information on the NCES and state department websites including published PLDs and released data available in data hubs was leveraged to conduct these comparisons.

⁴ Many of these individuals asked to remain anonymous for their participation in this process.

Evaluation Findings

The evaluation findings are outlined in this section for each phase of standard setting. Detailed findings by evaluation criteria are included in Appendix A. Included within the summary of each phase below are outstanding questions that remain given the lack of available information.

Design

The proposed plan for the standard setting (i.e., documentation of plan before execution phase) was not available for review so details about the standard setting process outlined in the technical report were evaluated as design elements.

The design phase of the 2024 standard setting processes did not include the CEQA who is responsibility it is to “determine the cut scores for performance levels on all statewide assessments” (70 OK Stat § 1210.508 (2024)). Given this responsibility, the lack of CEQA involvement would likely contribute to a disconnect between the policy goals and the standard setting results. The influence of policy on the standard setting design seemed limited to the PLDs which were provided as the frame of reference. However, the CEQA was also not involved in the development of the PLDs as specified in statute. There were content-based benchmarks provided as a reference for panelists but it was not clear if this was intended as a policy reference.

Methodologically, the 2024 standard setting was designed as a reasonably sound process to include experts from across the state, who would be trained to follow a sound method for making cut score recommendations focused on the performance expectations outlined in the PLDs with iterative opportunities to discuss and refine their judgments. However, the overarching policy influence was lacking from the design. Before the standard setting planning began, the state TAC advised SDE that a standard setting may not be necessary given the minor changes to the content standards and PLDs suggesting that the existing cut scores could be applied. This evaluation did include a review of the 2017 and 2024 PLDs (see detail in Appendix B) which showed that most of the 2017 expectations for Proficient were also included as expectations for Proficient in 2024. Further, the TAC suggested the state could consider equating the current cut scores to the 2024 OSTP or reviewing these cut scores through a standards validation process. Because the state proceeded with the standard setting activity, it suggested there was an intentional goal to set new standards but the policy underlying this was unclear. As a point of reference, the panelists were provided with content-based benchmarks established by SDE and Cognia but the intention behind providing these was not included in the report.

The following questions remain about the design phase.

- Why did the SDE decide to undertake the standard setting activity given the TAC advised that a full standard setting may not be necessary given the minor changes to the content standards and PLDs?
- Was there a specific policy goal or expectation for the impact of the recommended cut scores? Relatedly, what was the rationale to exclude impact data as a factor in the judgmental process?
- Was the standard setting plan reviewed by the state TAC or state policymakers? If so, how was their feedback incorporated into the design?
- What was the goal of providing content-based benchmarks as feedback during the standard setting process? Was this to encourage panelists to recommend cut scores in alignment of the expectations of the SDE and Cognia content experts?

Execution

The execution phase of the standard setting process was largely evaluated based on information available in the standard setting report.

The standard setting process began with training activities that provided panelists with the opportunity to understand the purpose of standard setting and the processes they would follow, practice the types of judgments they were to make, and build a shared understanding the frame of reference (i.e., PLDs). Because the standard setting plan documentation was not available, it is unclear if the final process deviated from the design or what the implication of any deviations would be.

Overall, panelists reported that they were comfortable with the process and felt the training sufficiently prepared them to participate in the process. Further, the panelists seemed to understand the feedback provided between rounds and were able to use this to refine their judgments in successive rounds. As noted in the design section, it was unclear how the content-based benchmarks were intended to influence the panelists' judgments and in the Execution phase it is unclear what type of influence they had.

The following questions remain about the execution phase.

- Was the standard setting conducted in a manner consistent with an approved plan?
- How were the results (and other information) presented to panelists between rounds? What interpretive guidance accompanied each type of information?
- What do the results by round (e.g., round 1 vs round 2 vs round 3, disaggregated by panelist) suggest about the influence of each type of feedback?

Outcomes

The evaluation components of the Outcomes phase are organized in two parts. The first is focused on the evaluation of the standard setting process results and the second is focused on the comparison of the impact of the standard setting results to a variety of other sources of information.

Standard Setting

The Outcomes phase of the standard setting process was evaluated based on information available in the standard setting report. From the standard setting, there are two sets of results – those that came from the individual panels and those that include the modifications suggested by the subset of panelists who participated in the vertical articulation review. The details in the standard setting report suggest that the panelists were able to develop familiarity with the PLDs and a shared understanding of expectations for students at each performance level and their recommended cut scores were relatively consistent within the panel. Further, the panelists felt confident in their recommended cut scores but had mixed perceptions of the impact (which was shown to them at the end). The vertical articulation panelists did express lower confidence in the overall results than the full panel indicated for their grade specific results.

The results of the standard setting (both from the panels and from the vertical articulation) were then presented to the CEQA for approval. However, the results were presented without the comparative context provided in the report. Figures 3 and 4 show the proficiency rates from the standard setting results (panel results and articulated results) along with the comparative values provided in the reports (content-based benchmarks and prior cut scores). Across all grades, the prior cut scores for ELA yielded

lower proficiency rates than any other cut scores. For mathematics, the proficiency rates were relatively consistent among the standard setting panel results, the vertical articulation results, and the content-based benchmarks (largest difference observed at grade 6) and all three produced higher proficiency rates than the prior cut scores.

Figure 3. ELA Proficiency Rates from Standard Setting Report

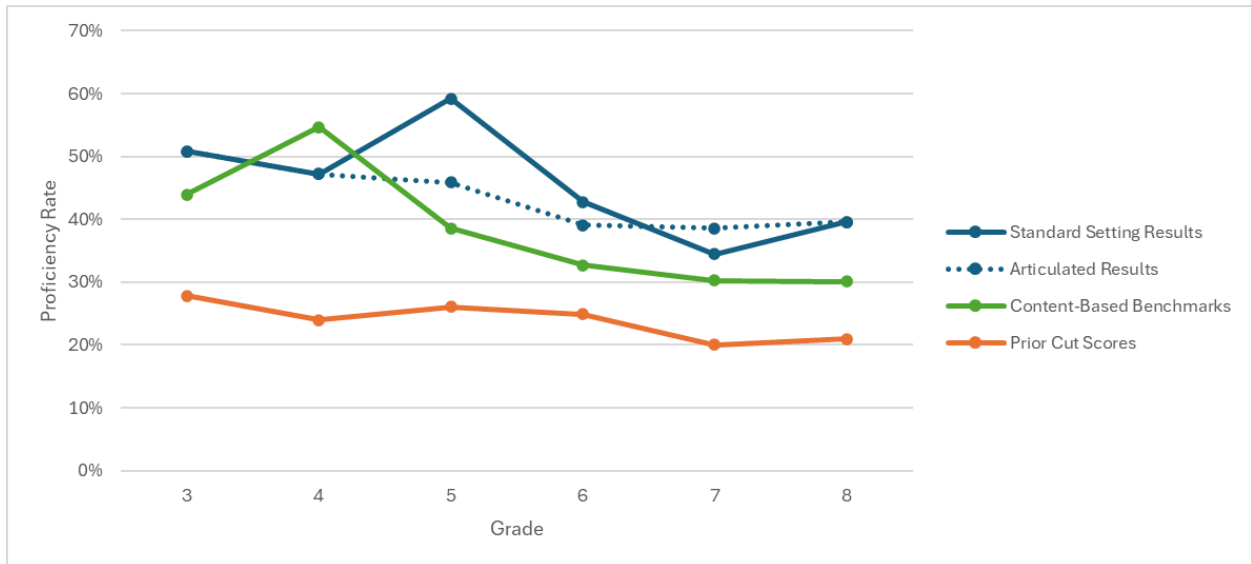
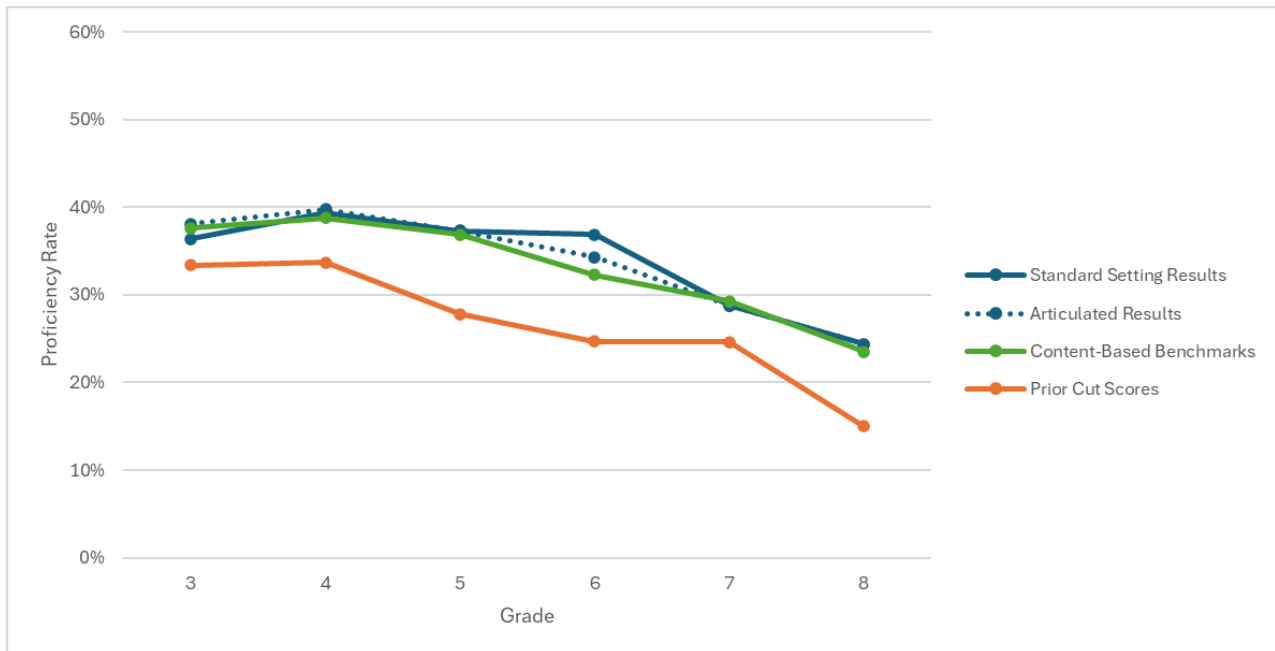


Figure 4. Mathematics Proficiency Rates from Standard Setting Report



The difference between the 2024 standard setting results and the prior cut score shown in Figures 3 and 4 were questioned by stakeholders (e.g., Martinez-Keel, 2024) who highlighted the lack of communication in advance of the score release (Oklahoma House of Representatives, 2024). These

questions were logical given the notable differences between the 2023 and 2024 proficiency rates. While it may have been understood that students did not experience this level of gain in one year, it was unclear to many stakeholders what the difference represented. Were the previous cut scores too high? Were the new cut scores too low? Which set of cuts represented Oklahoma’s educational assessment policy?

The standard setting process, resulting cut scores, and 2024 proficiency rates were discussed for months as state leaders debated who was responsible for the change that occurred from the previous cut scores (e.g., Martinez-Keel, 2025; Oklahoma State Department of Education, 2025).

The following questions remain about the outcomes of the standard setting.

- How were panelists instructed to evaluate the impact values presented at the end of the judgmental process? Were any benchmarks (e.g., impact based on NAEP, prior cut scores) provided for reference?
- Were the standard setting panelist evaluation results about the impact of the cut scores considered in the vertical articulation meeting?
- What communication efforts were made to explain the difference in impact between the 2024 cut scores and those from previous years (set in 2017) to stakeholders? (see Oklahoma State Department of Education, 2017 for example from previous standard setting)

Given the questions raised around the impact of the 2024 cut scores and Oklahoma statute §70-1210.508 which requires that the assessments should be comparable to those of similar states, the evaluation of the outcomes continued into investigating the 2024 proficiency rates against other metrics. The next section of this report compares the 2024 Oklahoma proficiency rates to those of other assessments of Oklahoma students, historical proficiency rates in Oklahoma, and assessments of students in neighboring states.

Comparison of Impact and Expectations

Comparing proficiency rates across assessments comes with some cautions as each assessment and group of students are unique. However, there are some commonalities that can be identified to facilitate comparisons:

- **Content Standards:** While they may be developed for each program, content standards for a given subject and grade will likely include notable similarities in terms of topics covered.
- **Performance Levels:** Each assessment program creates a system of performance levels by defining the numbers of levels and names of each level. However, each set of levels includes one which indicates an acceptable level of performance (e.g., “proficient”, “meets expectations”). Further, even though many programs identify an additional performance level above “proficient”, reporting the proficiency rate (percentage of students who achieve the proficient level or above) is a common way to identify what proportion of students meet the grade-level expectation.

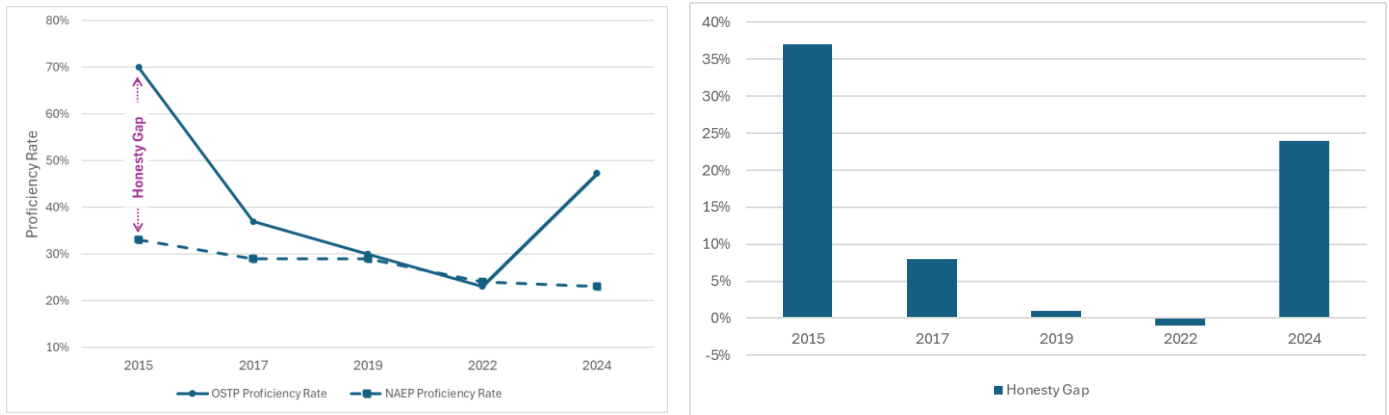
This evaluation further explored the comparison of Oklahoma proficiency rates to:

- those from the National Assessment of Educational Progress (NAEP) - 2024 and prior
- those of the states neighboring Oklahoma – state assessment and NAEP results

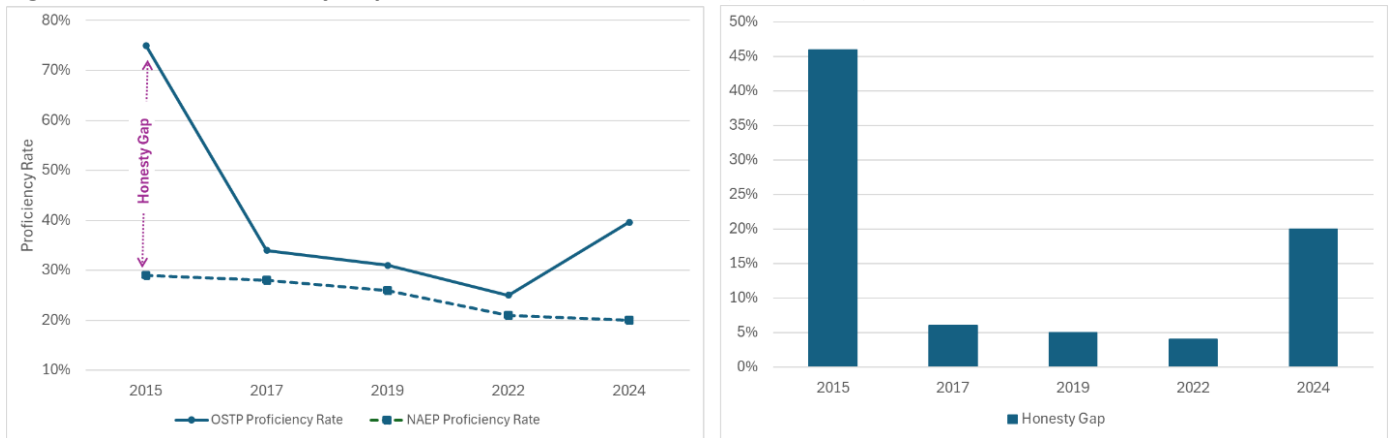
Comparison to the National Assessment of Educational Progress

In 2017, Oklahoma adopted the policy philosophy to align their assessments (and outcomes) with national assessments such as the National Assessment of Educational Progress (NAEP) and measures of college readiness including the ACT and SAT (Oklahoma Edge, 2017). Figures 5a and 5b show this concept of the Honesty Gap (described above) for Oklahoma ELA grade 4. The graph on the left shows the proficiency rates by reporting year and the graph on the right shows the exact size of the Honesty Gap. Similar comparisons are shown in Figure 6 for ELA grade 8 and in Figures 7 and 8 for mathematics grades 4 and 8, respectively. Although the amounts vary by subject and grade, the pattern across these four sets of comparisons show the impact of the 2017 policy goal to align with national standards such as NAEP which led to a reduction in the Honesty Gap, the continued impact of this goal through 2022, and the impact of the 2024 standard setting and resulting cut scores which led to a widening of the Honesty Gap.

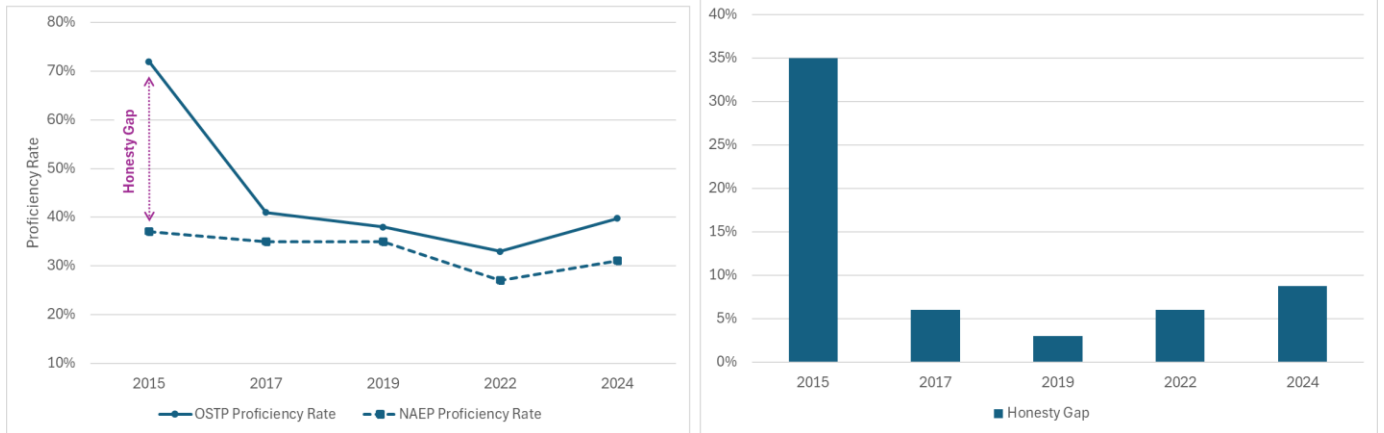
Figures 5a and 5b. Honesty Gap for Oklahoma ELA Grade 4 (2015-2024)



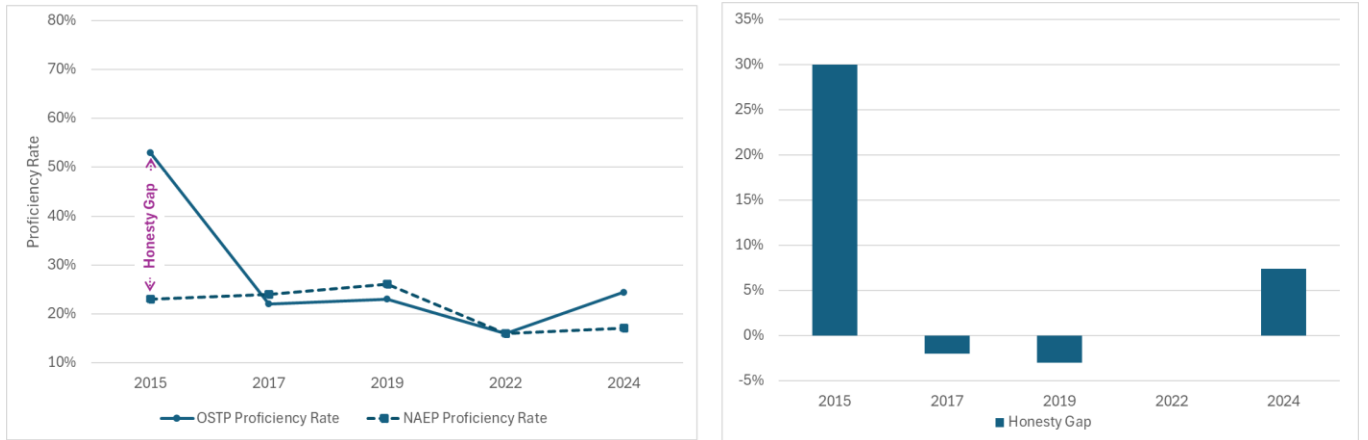
Figures 6a and 6b. Honesty Gap for Oklahoma ELA Grade 8 (2015-2024)



Figures 7a and 7b. Honesty Gap for Oklahoma Mathematics Grade 4 (2015-2024)



Figures 8a and 8b. Honesty Gap for Oklahoma Mathematics Grade 8 (2015-2024)



Oklahoma and Neighboring States

Figures 9 and 10 show the Oklahoma proficiency rates for ELA and mathematics based on the 2024 cut scores in comparison to the range of proficiency rates reported for those states neighboring Oklahoma (Arkansas, Colorado, Kansas, Missouri, New Mexico, Texas) for the same year (i.e., 2024), subject, and grade (see Appendix C for the detail behind these Figures). Although the exact results of this comparison vary by subject and grade, the overall perspective shows that Oklahoma’s proficiency rates are within this range, often close to the middle.

Figure 9. Oklahoma 2024 ELA Proficiency Rates Compared to Neighboring States

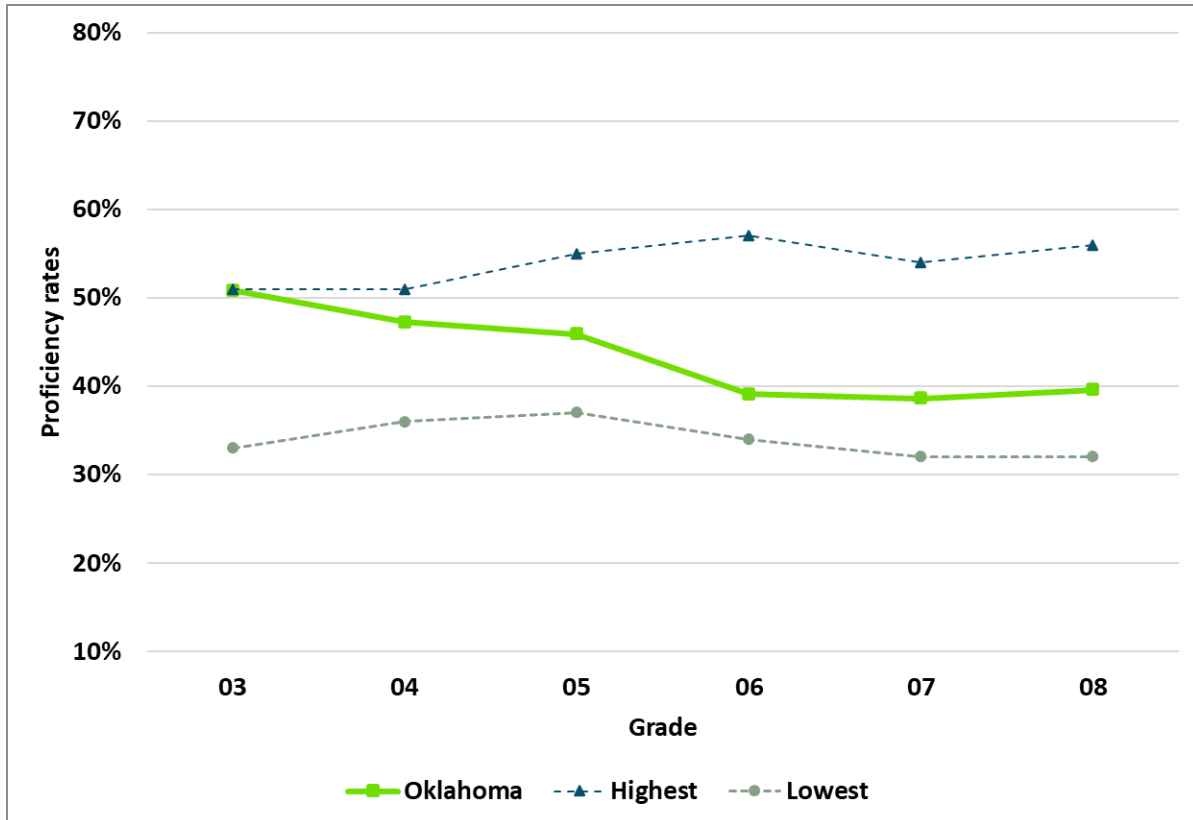
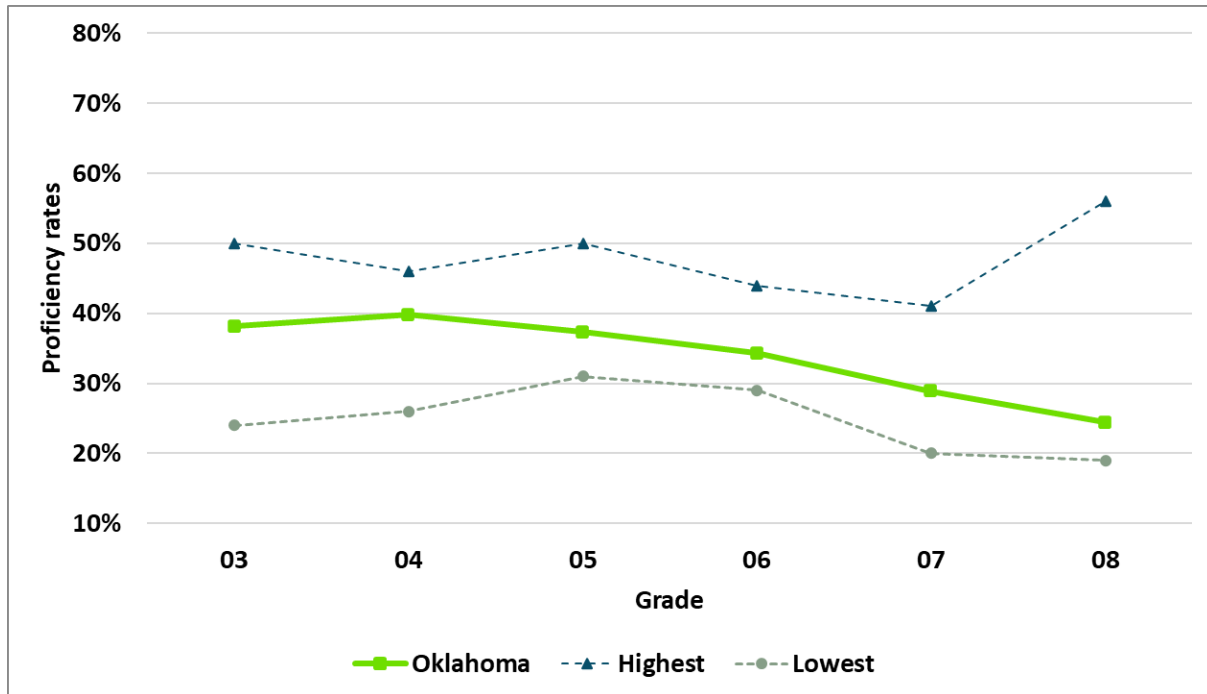


Figure 10. Oklahoma 2024 Mathematics Proficiency Rates Compared to Neighboring States



Oklahoma, Neighboring States, and NAEP

Figures 11 and 12 show the proficiency rates for ELA for this same group of states compared to the proficiency rates reported by NAEP in 2024 for grades 4 and 8, respectively. Similarly, the same results are presented for mathematics in Figures 13 and 14. The comparison of these two sets of results varies by state and grade with some showing larger differences and others showing more similar results. On average, states reported a proficiency rate 15% higher than that reported by NAEP for their state for ELA and 5% higher for mathematics. Oklahoma’s difference from NAEP was higher than this average for both ELA (average difference of 22% between grades 4 and 8) and mathematics (average difference 8% between grades 4 and 8). The difference between proficiency rates reported by states and those reported based on the NAEP assessment from within their state is what has been described in this report as the “Honesty Gap” (see Chu, 2025, Honesty Gap, n.d.,).

Figure 11. Comparison of 2024 ELA Grade 4 State and NAEP Proficiency Rates

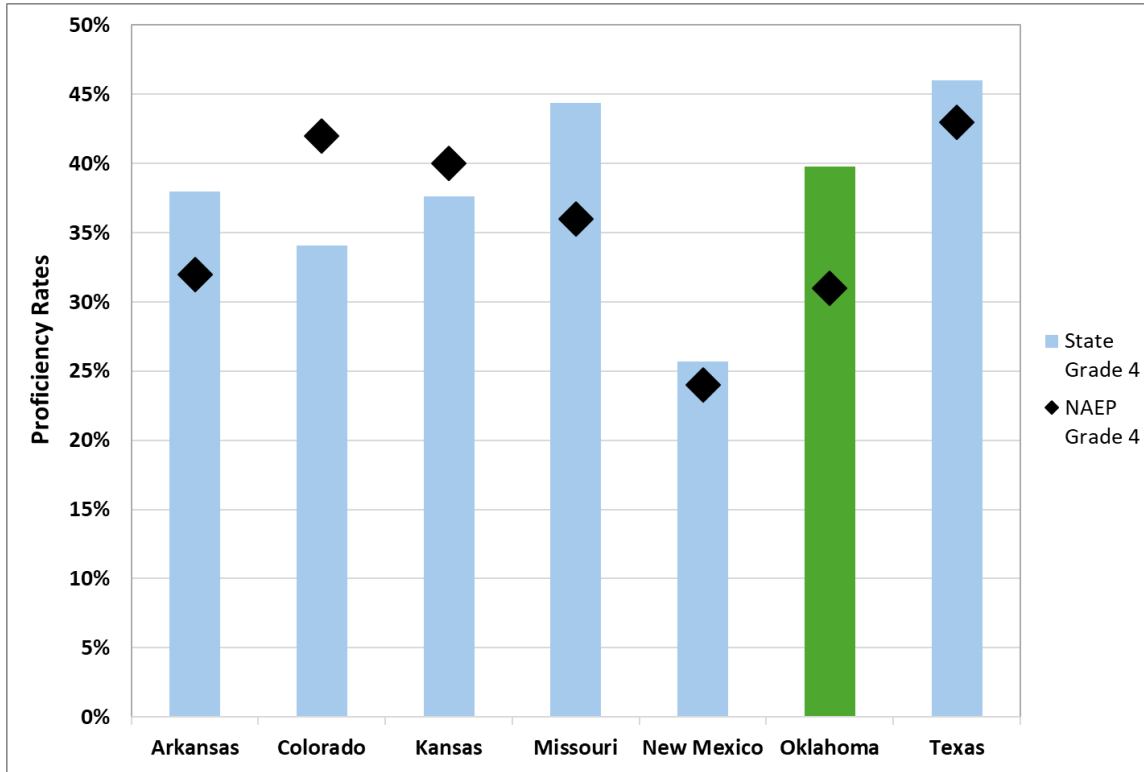


Figure 12. Comparison of 2024 ELA Grade 8 State and NAEP Proficiency Rates

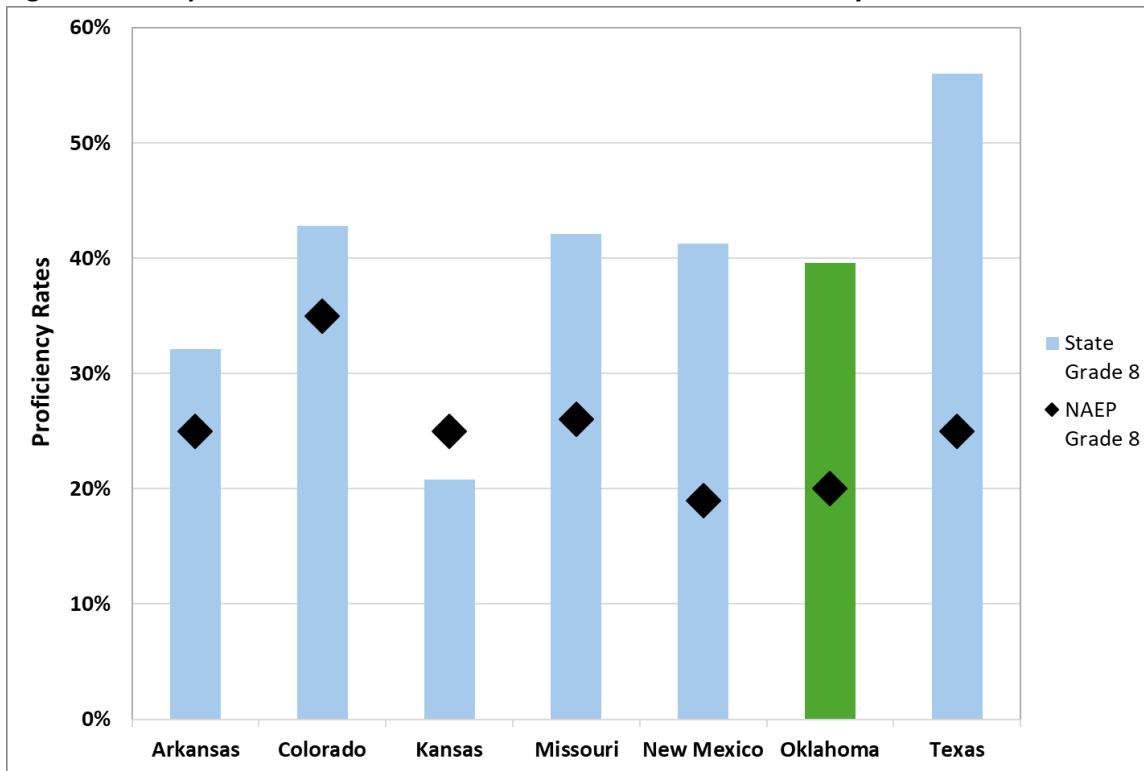


Figure 13. Comparison of 2024 Mathematics Grade 4 State and NAEP Proficiency Rates

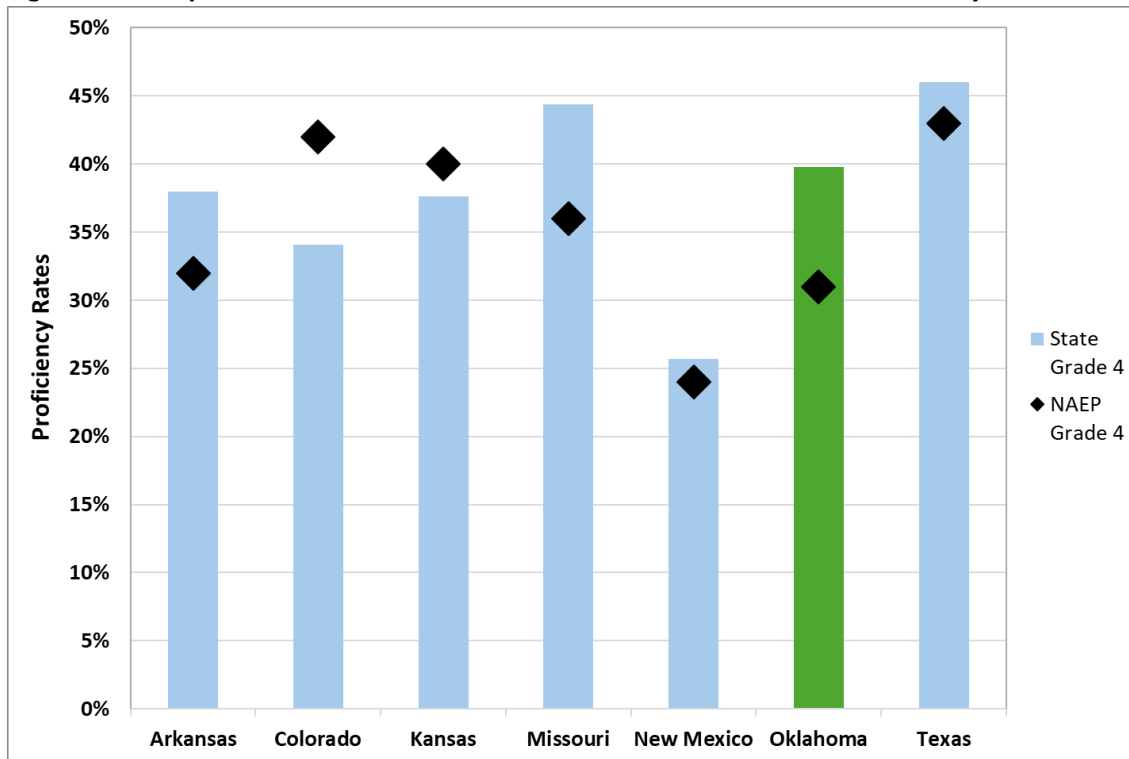
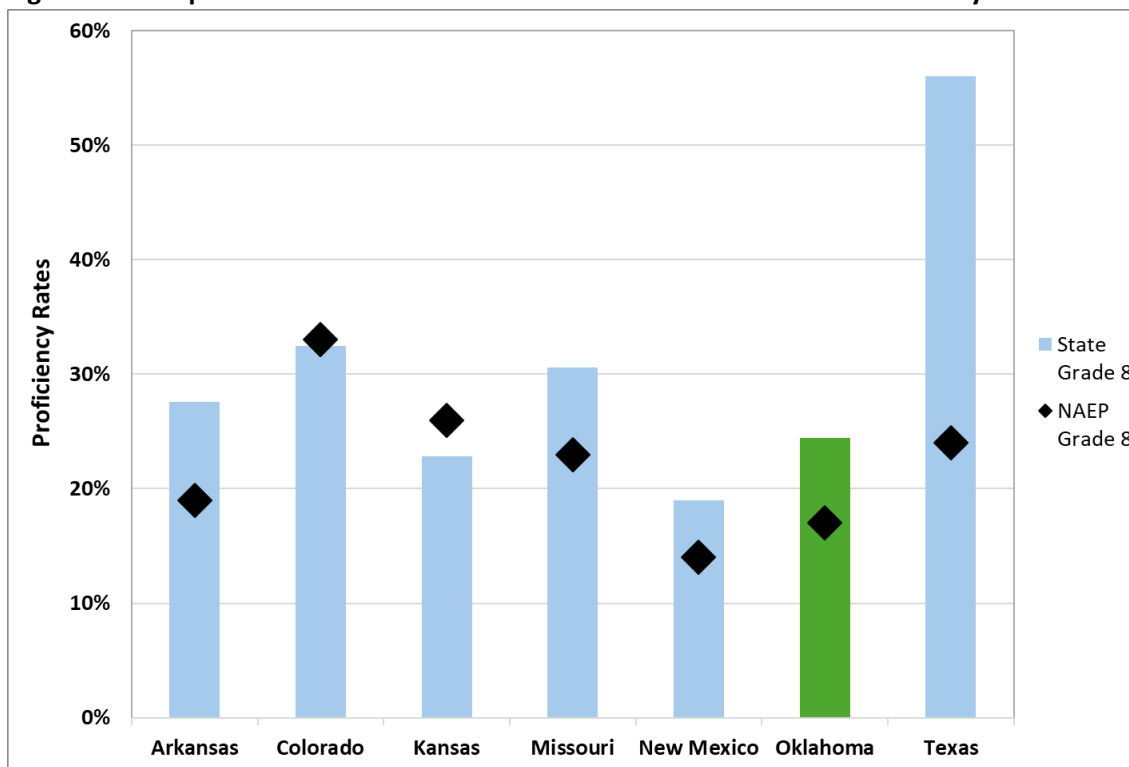


Figure 14. Comparison of 2024 Mathematics Grade 8 State and NAEP Proficiency Rates



Summary of 2024 Standard Setting Evaluation

This information gathered from the 2024 standard setting for the Oklahoma ELA and mathematics assessments in grades 3-8 was evaluated by considering the policy and psychometric elements of the design, execution, and outcomes phases of the standard setting process. The findings presented in this report reflect the information that was available for the evaluation including some documentation and input from those director or indirectly involved in the 2024 standard setting. Here we summarize the findings from our evaluation.

- **Design:** Because information was not provided about the planning for the standard setting (e.g., standard setting plan document, review notes from policy or psychometric groups), the design elements were inferred based on the standard setting report. In general, the standard setting design seemed reasonable including the formation of panels and selection of a standard setting method. However, there was no evidence as to how the design of the standard setting connected to the larger educational assessment policy other than the PLDs nor was the CEQA involved in this phase despite their responsibility for the cut scores. The standard setting did include content-based benchmarks established by Cognia and SDE but it was unclear the intention for including these and the expectation for how they would influence the results.
- **Execution:** The standard setting report included the results from the judgmental process but lacked some of the round-by-round results and information such as feedback presented to panelists, facilitate discussions, and how this influenced panelists' judgments. Overall, the panelists seemed comfortable with the standard setting process and had a general level of comfort with their results.
- **Outcomes:** The results appeared to represent the intended outcomes of the process: panelists made judgments about expected student performance in alignment with the PLDs which were translated into recommended cut scores. However, the results were presented for final approval without valuable contextual information. For example, how was the impact of the recommended cut scores presented and discussed? What historical and external evidence was provided as context to inform policy deliberations in determining the final cut scores?
- **Comparative results:** Comparisons of Oklahoma proficiency rates with NAEP and neighboring states on the 2024 assessments suggest that Oklahoma's proficiency rates are generally comparable to peer states and notably lower than the NAEP. Further, a historical review of the Oklahoma proficiency rates review shows the distinct influence of the 2017 policy goal to align with national expectations and the resulting reduction of the Honesty Gap which widened in 2024 based on the new cut scores.

As described in the standard setting report, many elements of the 2024 standard setting generally followed professional standards (AERA et al., 2014), particularly with respect to the *Execution* phase as described in this report. However, the lack of clear policy guiding this process and explaining the results raises key questions as to whether the results align with the intended interpretation of scores.

Next Steps and Recommendations

From the synthesis of the findings from the evaluation, we offer two recommendations in addition to some policy options for the CEQA to consider:

First, we recommend that potential adjustments to the cut scores from 2024 be informed by policy goals, external data, and vertical articulation data. Because standard setting is an integration of policy, content, and data, it is important to ensure that the system of performance standards and cut scores is aligned with other aspects of the assessment system. Whether adjustments should or should not be made is ultimately a policy decision that considers the available evidence shared in this evaluation.

Based on the findings in this report, the CEQA is encouraged to consider the following policy options with respect to applying the results of the 2024 standard setting to the 2025 OSTP administration.

OPTION A: Adjust 2024 cut scores to apply to the 2025 administration – Because alignment with policy goals is an important consideration, if the CEQA determines that there was a lack of policy alignment with outcomes of the 2024 standard setting, they could adjust the cut scores before applying them to the 2025 OSTP administration. The policy-guided nature of any adjustment should be made in such a way that it still ensures alignment with the stated performance expectations (i.e., PLDs) and based on empirical evidence (e.g., student performance data, benchmarks). It is important to note that making an adjustment to cut scores that apply to the 2025 OSTP administration has downstream effects on score reporting and the state's assessment vendor responsible for producing score reports as well as activities related to statewide accountability reporting.

OPTION B: Maintain the 2024 cut scores for the 2025 administration and implement any adjustments to cut scores for the 2026 administration – If there is not a clear policy goal against which to evaluate the cut scores or make adjustments for the current administration, the CEQA should reinforce its policy goals, communicate them to stakeholders, and use the results of this evaluation and the standard setting technical report to determine if adjustments to cut scores are necessary for the 2026 OSTP administration. This option provides an opportunity to coordinate on policy goals and would allow for adequate notice to the field regarding implementation of any changes.

OPTION C: Maintain 2024 cut scores with explanation of the policy rationale – As discussed in this report, our review of the available information about the 2024 standard setting suggests that it was conducted using an acceptable process. There are some areas of our evaluation where questions were raised that could not be answered with the available information. However, these questions do not necessarily represent concerns about the selection or execution of the standard setting methodology. Rather, our evaluation indicated that the strong policy perspective about alignment with national standards that influenced the 2017 cut scores did not seem to be a consideration in the 2024 standard setting process. This observation suggests a policy shift; however, this shift was not intended or recommended by the CEQA. Alternate policy interpretations may be reasonable but would benefit from clear communications about the frame of reference and rationale for each perspective and be determined by the CEQA as the policy body responsible for setting the cut scores.

Second, future standard setting processes must involve the CEQA in their legislatively authorized roles, as well as those with expertise in areas related to standard setting, to provide guidance regarding the design, execution, and outcomes. A key aspect of the validity argument supporting the use of assessment results is the alignment of components within the larger assessment system. With respect to standard setting, the cut scores should reflect the intended interpretation of assessment results as outlined in the performance level descriptors (PLDs). Given the need for these elements to be aligned, it is logical that the CEQA is legislatively responsible for both the PLDs (narrative descriptions of student performance expectations) and the cut scores (empirical representations of student performance expectations). To ensure this alignment, the CEQA should be involved throughout the standard setting processes (either setting new cut scores or validating existing cut scores) as follows.

- **PLD Development:** The CEQA should provide leadership in the development of the range PLDs from the policy PLDs (established through legislation). More specifically, the CEQA can identify relevant benchmarks such as alignment with national achievement indicators and references to how expectations can be operationalized across grades and subjects as well as identify those who should be involved in creating the range PLDs. The CEQA should collaborate with those with expertise in these activities to design and lead workshops following the policy guidance provided.
- **Standard Setting Design:** The CEQA should be directly involved in the design of the standard setting activities in collaboration with the standard setting experts/study leaders and the Oklahoma State Board of Education (SBE). More specifically, the CEQA should be involved in identifying the frame of reference and benchmarks that will help guide the panelists' judgments, creating a recruitment strategy for expert panelists, and developing the overall process for how feedback information will be used to inform panelist judgments.
- **Standard Setting Execution:** The CEQA should observe the execution of the standard setting to ensure the process is executed as outlined in the agreed-upon design, evaluate any deviations from the design, and be available to discuss any policy-related questions that arise with the standard setting leaders.
- **Standard Setting Outcomes:** The CEQA should continue their role of approving the cut scores but with additional knowledge and understanding of how their policy goals informed all aspects of standard setting, starting from the PLD development process, the design of the standard setting, the execution of the standard setting process, identification of relevant benchmarks against which the impact of the standard setting results can be reviewed, and the vertical articulation of the results to inform a coherent system of cut scores across grades and subject areas. The benchmark data that standard setting vendors often include in presentations to policymakers are the range of recommendations from panelists, historical student performance, current student performance, national indicators, and process evaluation from panelists that communicate their experiences with the process.

Given the technical nature of these topics and that the CEQA members may not have experience working with the standard setting process, it is further advised that the CEQA members receive resources and training available to prepare for their oversight role in these areas along with technical guidance on the interpretation on the study results.

References

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- Assessment HQ. (n.d.). Why assessments: Honesty gap. Assessment HQ. <https://assessmenthq.org/why-assessments/honesty-gap/>
- Chu, D. (2025, February 20). Mind the honesty gap. Fordham Institute. <https://fordhaminstitute.org/national/commentary/mind-honesty-gap>
- Cizek, G. J. (Ed.). (2012). *Setting performance standards: Foundations, methods, and innovations (2nd ed.)*. Routledge. <https://doi.org/10.4324/9780203848203>
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Sage Publications Ltd.
- Ferrara, S., Davis-Becker, S., Kannan, P., & Reynolds, K. (in press). *Standard Setting: A Cognitive and Social Model*. In M. Pitoniak & L. Cook (Eds.), *Educational Measurement (5th ed., in press)*. Oxford Press.
- Honesty Gap. (n.d.). What is the honesty gap? Honesty Gap. <https://honestygap.org/what-is-the-honesty-gap/#:~:text=Frequently%2C%20states'%20testing%20and%20reporting,as%20an%20%E2%80%9Chonesty%20gap.%E2%80%9D>
- Martinez-Keel, N. (2017, October 9). Oklahoma changes test score rates, prepares for shock to the system. *The Oklahoman*.
- Martinez-Keel. (2024, August 21). Oklahoma's education department quietly made changes that will cause big jumps in state test scores. *Oklahoma Voice*, <https://oklahomavoice.com/2024/08/21/oklahomas-education-department-quietly-made-changes-that-will-cause-big-jumps-in-state-test-scores/>
- Martinez-Keel, N. (2025, March 18). Benchmarks as influential: When setting test rigor, Oklahoma education officials wanted to avoid a 'thumb on the scale.' *Oklahoma Voice*. <https://oklahomavoice.com/2025/03/18/when-setting-test-rigor-oklahoma-education-officials-wanted-to-avoid-a-thumb-on-the-scale/>
- Oklahoma House of Representatives. (2024, August 22). McBride, Dick Lowe Discuss Student Test Scores. *Oklahoma House of Representatives*. https://www.okhouse.gov/posts/news-20240822_1
- Oklahoma Legislature. (2024). *Statewide student assessment system*. Oklahoma Statute § 70-1210.508.
- Oklahoma State Department of Education. (2025, March 6). Premature cut score release undermines Oklahoma's education [Press release]. Oklahoma State Department of Education. <https://oklahoma.gov/education/newsroom/2025/march/premature-cut-score-release-undermines-oklahomas-education.html#:~:text=OKLAHOMA%20CITY%2C%20OK%20%E2%80%93%20Under%20the%20direction%20of,serious%20concerns%20about%20political%20interference%20in%20public%20education.>

Oklahoma State Department of Education (2017). Giving Oklahoma Students a Competitive Edge: FAQs for Legislators / Test Results 2017. Oklahoma City, OK: Author.

Virginia Department of Education (2022). *Our Commitment to Virginians: Higher Expectations and Excellence for All Students*. Office of the Governor of Virginia.

<https://www.education.virginia.gov/media/governorviriniagov/secretary-of-education/pdf/Our-Commitment-to-Virginians.pdf>

Appendix A: Evaluation Questions, Information, and Findings by Phase

The tables in this appendix provide detail on the specific evaluation questions identified for each phase, the information available to address the question, and the findings from the review of the available information.

Table A.1 Design Phase Evaluation Questions, Information, and Findings

Evaluation Questions	Information Available	Evaluation Findings
<p>What educational assessment policy goals informed the design of the standard setting and selection of the final cut scores?</p>	<p>Based on conversation with a former TAC member, the SDE was advised they could maintain the current cut scores on the OSTP or review them through a standards validation process. Based on conversation with SDE staff, the goal of the 2024 standard setting was to establish cut scores that represented the opinions of the educators (i.e., expert panelists). Evidence of how policy considerations were defined and incorporated into the standard setting plan document was not provided.</p>	<p>Information was not available to clearly delineate the policy goal for conducting the standard setting (in lieu of an equating or standards validation). The CEQA was not involved in setting policy goals for the standard setting (e.g., purpose for standard setting, identifying benchmarks).</p>
<p>How did the design operationalize the policy goals as a frame of reference for the standard setting panelists?</p>	<p>The state-level policy PLDs were operationalized through the development of range PLDs in a separate activity by detailing expectations for students in each subject and at each grade in alignment with the content standards. Standard setting panelists used these range PLDs as the basis for their judgements in the standard setting process. The training materials emphasized “content-based benchmarks” that were developed by the state content experts and their assessment vendor.</p>	<p>Based on the information available, the policy influence in the design was limited to the policy PLDs which were operationalized as range PLDs and the primary focus of the panelist judgments. The CEQA was not involved in the design phase to determine how the policy goals would be shared with the panelists.</p>
<p>How was the panelist recruitment strategy designed and executed?</p>	<p>The standard setting technical report describes goals for panel recruitment and demographic information about each panel.</p>	<p>The panelist selection seemed reasonable and in line with professional standards as panelists represented a range of experience and district type. Evaluation comments suggested at least one panelist was assigned to a grade level where they did not have experience and indicated concern about including panelists who did not have recent classroom experience. However, this did not appear to be pervasive across panels.</p>

Evaluation Questions	Information Available	Evaluation Findings
Was the selected standard setting method reasonable given the design of the assessment, the experts selected to make judgments, and the intended interpretation of scores?	ID matching method (see Ferrara et al., in press) was implemented using a specific statistical criterion (RP67) for ordering items. Panelists were presented with a set of items (ordered item booklets, OIBs) that met the blueprint.	The choice of method was reasonable given the design of the assessment, the expert panelists selected to participate in the process, and the empirical information available to inform the process.
What empirical information was identified to be included in the standard setting process?	The standard setting process, by design, included information about item difficulty as the items were presented in order (easiest to most difficulty). At the end of the judgmental process, panelists were able to view the impact of their standard setting judgments. Only those selected panelists who participated in the vertical articulation were able to consider the impact in their recommended articulation judgments.	Item difficulty information was appropriately incorporated into the design of the OIB as specified in the method. The study design did not allow panelists to consider the impact except in the vertical articulation process which was focused on the pattern of impact across grades within each subject. Given that the goal of vertical articulation is to ensure a consistent pattern of results, the impact may only have been evaluated across grades (not within grades).
What benchmarks were identified to be included in the standard setting process?	During the judgmental process, panelists were provided with content-based benchmarks that were established by SDE and Cognia before the standard setting.	It is unclear how the content-based benchmarks aligned with the design of the study or how they were intended to influence the panelists' judgments.
Was the standard setting process designed in a way that expert panelists could meaningfully contribute to the cut score recommendations?	The standard setting vendor designed a process that began with training on the standard setting method, an opportunity for panelists to orient themselves to the range PLDs as a frame of reference, and then an iterative judgmental process where they could receive feedback and refine their judgments.	The design seems appropriate for the panelists with the inclusion of training activities, iterative judgment process, and opportunities to collect feedback information.
Were the relevant stakeholder groups involved in the design of the standard setting process?	Oklahoma legislation mandates responsibility for the academic content standards to the State Board of Education, assessments to SDE with support of technical experts in assessment, and the cut scores to the CEQA.	Despite their responsibility for the final cut scores, the CEQA was not involved in the design of the process and was only presented with a final set of recommendations from the standard setting. Further, there was no information available as to whether technical experts (e.g., the TAC) reviewed the standard setting plan documentation or if their feedback was incorporated.

Table A.2 Execution Phase Evaluation Questions, Information, and Findings

Evaluation Questions	Information Available	Evaluation Findings
Was the standard setting process executed following the design as outlined in the standard setting plan?	Documentation on the standard setting plan documentation was not available for review.	No information about the plan was available to be able to determine whether it was executed as expected.
Were the training activities appropriate to prepare the panelists to participate in the standard setting process?	<p>The training activities covered the assessment, the standard setting process, practice with the standard setting method and judgmental task, and familiarity with the feedback format.</p> <p>Before each round, panelists were asked to complete a survey indicating how ready they felt to engage in the next step in the process. At the end of the standard setting meeting, panelists completed a final evaluation of the entire process including the training.</p>	<p>The training activities covered the key topics required to prepare standard setting panelists for their activities.</p> <p>The standard setting report did not include results of the surveys conducted before each round, but the final evaluation results indicated panelists felt prepared to make the types of judgments required in the process.</p>
Did the panelists follow the judgmental task as prescribed in the standard setting method?	The standard setting report did not provide round-by-round judgmental results but indicated that each round of results was reviewed for aberrance. No ratings were identified as aberrant. Further, panelists responded to questions in the final evaluation about their understanding of the judgmental task.	Panelists indicated that they were comfortable with the judgmental task and the lack of aberrance suggested the panelists followed the task as prescribed.
Did the panelists understand the feedback information provided and appropriately integrate this information into their iterative judgmental process?	<p>Panelists were provided with feedback after each round that included their recommended cut scores, panel results, and the content-based benchmarks. The facilitator led the panelists in a discussion of these results and instruction as to how they should integrate them within their next round of judgments.</p> <p>The standard setting report only includes item-level results by round (not the panelist or panel level results) which did not allow an evaluation of the impact of each type of feedback. However, the workshop evaluation included questions about understanding feedback and how to use it in later rounds of judgments.</p>	<p>Panelists indicated via the evaluation that they understood how to use the feedback in their judgments.</p> <p>The standard setting report notes that panelists were told the content-based benchmarks “were provided for their consideration, and not to influence their judgments” but it is unclear how this information was intended to be used or what type of influence these benchmarks had. One ELA evaluation comment suggested a panelist felt pressured to align with expectations but this did not appear to be the sentiment of the full panel.</p>

Table A.3 Outcomes Phase Evaluation Questions, Information, and Findings

Evaluation Questions	Information Available	Evaluation Findings
Did the panelists develop a shared understanding of performance level expectations?	Panelists reviewed and discussed the PLDs as part of the training activities, practiced using the PLDs to make judgments, and discussed their judgments after each round to inform their judgments in the successive round.	The time allocated to the review and discussion of the PLDs seemed reasonable and panelists overall rated this part of the training as successful.
Were the panelists' judgments consistent with the item difficulty?	The standard setting report includes item-level judgments for each panel (not by panelist) by round. Given the lack of detailed information, only visual inspection of the results was possible which suggests that panelist judgments became more consistent with the item difficulty across the rounds of the process.	Overall, the level of consistency between panelists' final judgments and item difficulty appeared reasonable.
Were the panelists' final recommendations consistent , indicating they had similar expectations for students at each performance level?	The standard setting report includes the standard error of the final recommendations.	Overall, the reported standard errors indicate that the panelists were reasonably consistent in their final recommendations.
Were panelists confident in the results from the standard setting study?	<p>Panelists responded to evaluation questions about their confidence in the results and the resulting impact.</p> <p>For those panelists who participated in vertical articulation, panelists responded to an evaluation from that activity about their perception of the process, their judgments, and the results.</p>	<p>Generally, the panelists expressed confidence in the results from the standard setting but had mixed perceptions of the impact. The ELA panelists in most grades indicated that there should be fewer students in the <i>Below Basic</i> category and more students in the <i>Basic</i> category. In most grades, the mathematics panelists indicated that there should be fewer students in the <i>Below Basic</i> and <i>Basic</i> categories and more students in the <i>Proficient</i> and <i>Advanced</i> categories.</p> <p>In the vertical articulation, the panelists overall indicated lower confidence in the results (compared to the standard setting panels) with some comments indicating a lack of comfort making judgments about grades they do not work with.</p>

Evaluation Questions	Information Available	Evaluation Findings
<p>Were the results reasonable based on relevant internal and external benchmarks?</p>	<p>The standard setting results were reviewed vertically (across grades) by panelists representing each panel within each subject. Standard setting leaders adjusted the recommended cut scores based on feedback from this panel.</p> <p>The standard setting report documents the impact of the panel recommendations, the content-based benchmarks, the prior cut scores, and the most recent NAEP results as points of comparison.</p>	<p>The articulated results represent the expectations set by the larger standard setting panel, smoothed through the vertical articulation process.</p> <p>Although comparative impact values (e.g., NAEP, prior cut scores) are provided in the report, there is no evidence that these were included in the standard setting meeting, vertical articulation discussions, or presentation to the CEQA as benchmarks. Nor does the report indicate if these benchmarks were provided as an indication of “reasonable” impact.</p>

Appendix B: Oklahoma PLD Comparisons between 2017 and 2024

Oklahoma’s 2017 PLDs were compared to the 2024 PLDs to see if there were any significant differences in performance expectations between the two. These comparisons described below were conducted by individuals who have worked with several sets of PLDs across diverse subject areas. In some cases, the differences between the two sets of PLDs were based on a change in format as the 2024 PLDs were much more detailed compared to the 2017 PLDs which were more text-based descriptions of key expectations sampled from across the grade level standards. Therefore, the 2024 PLDs appear to have a greater number of expectations for proficiency. However, more detail does not necessarily mean higher expectations. The intention of this review was to attempt to match the content of each expectation (e.g., main idea) within the 2017 PLDs and determining if there appeared to be a match in expectations, raised expectations, or lowered expectations.

The review here is focused on the expectations for the Proficient level. In terms of the other performance levels, 2017 uses "Limited Knowledge," and "Unsatisfactory" to describe the two lowest levels, whereas 2024 uses "Basic" and "Below Basic"; both PLDs describe partial or inconsistent mastery at these lower levels, but the 2024 Basic level descriptions include more specific foundational skills such as recognizing simple sentences and basic punctuation, suggesting a clearer developmental progression. The Advanced level in both years describe superior performance with consistent, skillful application of all proficient skills plus additional depth

English Language Arts (ELA)

The comparison of the expectations in the Oklahoma ELA PLDs between 2017 and 2024 across grades 3 to 6 (grades 7 and 8 PLDs were not available for 2017 in English Language Arts) is shown in Table B.1. Overall, these results show that almost all of the 2017 expectations for proficiency were found within the 2024 expectations. The minor exceptions are three expectations (across grades) where there was not a content match in the 2024 PLDs and then one where the expectation for proficiency appeared to be higher in 2024. Grade level summaries that highlight the key similarities as well as areas where the differences were identified are included below the table.

Table B.1 Comparison of Expectations in Oklahoma ELA Proficient Descriptors: 2017 to 2024

Grade	Match	Lowered expectations in 2024	Raised expectations in 2024	No content match identified
3	14	0	0	0
4	13	0	0	1
5	8	0	1	1
6	13	0	0	1
7	<i>2017 PLDs not available for review</i>			
8	<i>2017 PLDs not available for review</i>			

Grade 3: In both years, proficient students are expected to identify main ideas and demonstrate skills in organized writing. Additional expectations are detailed in the 2024 PLDs, where the demonstration of the following knowledge and skills is explicitly stated: more emphasis on structured writing (paragraphs), use of transitions, and expanded vocabulary work.

Grade 4: Similar core skills are stressed in both years including comparing genres, the recursive writing process, and identifying/understanding literary elements. The 2017 PLDs include an expectation for students to recognize the paraphrase of an original text.

Grade 5: Both PLDs require proficient students to demonstrate understanding of literary elements, genre recognition, and organized writing. However, the 2024 PLDs place more emphasis on summarizing a story and connecting the importance of details to a main idea. The 2017 PLDs include an expectation for students to recognize the paraphrase of an original text.

Grade 6: Both PLDs focus on partial to full mastery of comprehension, paraphrasing, and writing processes, with command of grammar and research skills. The 2024 PLDs build on these expectations at this grade-level with explicit emphasis on analyzing perspectives, literary devices, and formal essay composition, including thesis development and citation styles.

Mathematics

Overall, Oklahoma’s expectations for Proficient students in 2024 are largely similar to those in 2017 for mathematics in terms of core content areas such as number operations, fractions and decimals, geometry, measurement, data analysis, and problem-solving. Similar to ELA, the 2024 PLDs tend to be more detailed and explicit about certain skills and processes.

The comparison of the expectations in the Oklahoma Mathematics PLDs between 2017 and 2024 across grades 3 to 8 PLDs is shown in Table B.2. Overall, these results show that most of the 2017 expectations for proficiency were matched to the 2024 expectations. At each grade level, there were 1-4 expectations that appeared lower in 2024 and 0-3 expectations that appeared higher in 2024. Grade level summaries that highlight the key similarities as well as identified differences are provided below.

Table B.2 Comparison of Expectations in Oklahoma ELA Proficient Descriptors: 2017 to 2024

Grade	Match	Lowered expectations in 2024	Raised expectations in 2024	No content match identified
3	6	1	3	2
4	7	1	2	1
5	8	1	1	1
6	10	2	0	2
7	4	4	1	1
8	5	2	1	1

Grade 3: The PLDs from both years expect mastery of grade-level skills including comparing and ordering whole numbers, solving addition/subtraction/multiplication problems, understanding fractions, patterns, angles, measurement, and data analysis for students at this level. In 2024, there appear to be some higher expectations around rules for patterns, using data, and three-dimensional figures. In contrast there are lower expectations around working with fractions. The 2017 PLD includes concepts such as understanding the value of coins (which is not in the 2024 PLD) and the 2024 PLDs include concepts such as rounding and using multiplication and division facts (which are not in the 2017 PLD).

Grade 4: The PLDs from both years expect mastery of fractions, decimals, multi-step problems, geometry (quadrilaterals, polygons), measurement, and data representation for students at this level. In addition, for students at this level, the 2024 descriptor includes higher expectations around the value of money and working with data. In contrast, the 2017 PLD included higher expectations for estimation.

Grade 5: The PLDs from both years expect mastery of fractions, decimals, division with remainders for fractions/decimals, coordinate plane graphing, expressions, volume/perimeter, angle comparison, and data analysis and measurement. The 2024 descriptor add additional emphasis on analyzing graphs whereas the 2017 PLD included higher expectations on comparing representations of values.

Grade 6: The PLDs from both years expect mastery of integers, rational numbers, exponents, ratios, unit rates, expressions/equations/inequalities, transformations, symmetry, area calculations, angle relationships, probability, and data interpretation for students at this level. In contrast to the 2024 PLD, the 2017 PLD has slightly higher expectations around measures of central tendency and explaining the outcomes of experiments.

Grade 7: The PLDs from both years expect mastery of rational numbers, proportional relationships, equations/inequalities, surface area/volume, circle measurements, transformations, central tendency, and probability for students at this level. The 2024 descriptor add expectations around graphing the effect of dilations, translations, and reflections, but lowered expectations around equations, inequalities, and expressions.

Grade 8: The PLDs from both years expect mastery of scientific notation, real/irrational numbers, linear functions and equations, Pythagorean theorem, surface area/volume, scatterplots, and probability for students at this level. The 2024 descriptor included raised expectations around the relationships between independent and dependent variables. In contrast, the 2024 descriptor included lower expectations around linear functions and equations and interpretation of scatterplots.

Appendix D: Proficiency Rates for Neighboring States

The information provided in this appendix was used to create Figures 9 and 10 in the report which defines the range of proficiency rates for the states neighboring Oklahoma.

Table D.1 ELA Proficiency Rates for Oklahoma and Neighboring States (detail of Figure 9)

Grade	Arkansas	Colorado	Kansas	Missouri	New		Texas	Highest	Lowest
					Mexico	Oklahoma			
3	36%	42%	38%	42%	33%	51%	48%	51%	33%
4	36%	42%	43%	46%	42%	47%	51%	51%	36%
5	37%	47%	40%	43%	41%	46%	55%	55%	37%
6	34%	44%	34%	38%	40%	39%	57%	57%	34%
7	32%	46%	32%	38%	41%	39%	54%	54%	32%
8	32%	43%	21%	42%	41%	40%	56%	56%	32%

Table D.2 Mathematics Proficiency Rates for Oklahoma and Neighboring States (detail of Figure 10)

Grade	Arkansas	Colorado	Kansas	Missouri	New		Texas	Highest	Lowest
					Mexico	Oklahoma			
3	38%	42%	50%	41%	24%	38%	42%	50%	24%
4	38%	34%	38%	44%	26%	40%	46%	46%	26%
5	38%	37%	31%	38%	32%	37%	50%	50%	31%
6	38%	29%	30%	44%	30%	34%	39%	44%	29%
7	38%	30%	28%	41%	20%	29%	34%	41%	20%
8	28%	33%	23%	31%	19%	24%	56%	56%	19%