

Writing Scoring Workshop

How is Artificial Intelligence (AI) trained for automated scoring?

The AI is trained by examining about 2,000 to 2,500 student writing samples in each scoring category (1-4 for Purpose, Focus, & Organization and Evidence & Explanation, 0-2 for Conventions) at each grade level. This allows the machine to recognize the language and writing patterns for each scoring category. The machine looks specifically for language and writing patterns at each grade level. Even though the rubrics are the same in grades 3-5, 6-8, and 9-12, student writing is scored in comparison to the expectations for only that grade level. Conventions are slightly different, and are based primarily on usage, spelling, etc.

How or why would a student response be routed for hand scoring?

For the summative assessment, about 500 responses are randomly selected for hand scoring. Additionally, in both the summative and interim assessments, a student response will be routed for hand scoring if it is flagged with one of the following condition codes:

- **No Response** - The response was empty or consisted only of white space (space characters, tab characters, return characters).
- **Not Enough Data** - The response has too few words to be considered a valid attempt at the prompt.
- **Duplicate Text** - The response contains a significant amount of duplicate or repeated text.
- **Prompt Copy Match** - The response consists primarily of text from the passage.
- **Unusual Vocabulary** - Most words in the response do not appear in typical responses.
- **Non-Specific** - The response displays characteristics of condition codes assigned by humans that do not align with the above condition code categories.

How does the scoring of the writing assessment differ from teachers' grading of papers?

Teachers know their students; the AI and third-party hand scorers do not. When teachers grade papers with a rubric, they already possess knowledge about what has been taught and what the student is able to do. A teacher might start grading at a proficient level and be looking for evidence of mistakes that might lower the score, or start looking for evidence of exceptional writing that could result in advanced scores. The scoring engine and hand scorers do not operate like that. They are starting from zero and looking for evidence of knowledge and mastery.

Remember, in the classroom, teachers grade student writing to provide feedback for learning and improvement. On the assessments, writing items are scored, not graded. The AI or human scorers are simply assigning point values to samples of student writing. These scores are not meant as a learning tool, rather as a snapshot of a student's current performance against the standards.

How is Lexile used on the WY-TOPP assessment?

The Lexile® score provided on a student's Individual Student Report is a result of a conversion of the

summative ELA assessment scale score to a Lexile® measure. A linking study for ELA and mathematics took place at the end of June 2018 to determine final conversions.

Lexile® is a quantitative measure used to select texts for the ELA assessment; however, content experts also make an overall judgment about each passage based on a combination of both qualitative and quantitative features. The table below shows the general Lexile® ranges for each grade:

Grade	Lexile
Third	450-900
Fifth	770-1050
Seventh	955-1200
Ninth	1080-1400

Do students need to use evidence from all sources?

No. Evidence should support an opinion/idea/claim a student is stating. They must provide at least two supporting evidence examples within the essay but students do not have to refer to every source provided. In the higher grades, students are expected to address counterarguments. Scores for Evidence and Elaboration could be impacted if students do not support their ideas including evidence from counterargument.

Is formatting considered on the WY-TOPP writing prompt?

Formatting, as a general rule, is ignored. For example, students don't have to use formal letter structure when asked to "write a letter." "Multi-paragraph" doesn't really mean the student pressed the "enter" key more than once or indented. Rather, they included multiple ideas that are considered separate, but which are connected throughout the entire writing piece.

Is spell check provided on WY-TOPP?

Yes. There is a basic spell check built into the Secure Browser. Students should be encouraged and reminded before testing starts to utilize all available Universal Supports. More information about the Universal Supports available to all students can be found in the [Wyoming Accessibility and Accommodations Guide](#).

How does the writing portion impact a student's overall ELA score?

Scoring is based on an algorithm, so there is not a set percentage, but the overall ELA score is, on average, about 19% impacted by writing, with the other 81 percent of the score being influenced by the students' performance on the non-writing items. Statistical models adjust for difficulty, so students who receive a "difficult" prompt are not penalized.

My student's score is "grayed out" or came back with a score of 0; what happened?

If scores come back "grayed out," teachers need to know the writing sample has been routed to additional scoring based on a condition code – most likely due to too much copied text from the prompt or too much duplicated text. About 70% of similar/identical text will trigger this condition code.

What is the appropriate usage of WY-TOPP writing data?

As with all statewide assessment scores, the scores and rubric data provided to teachers through the

Centralized Reporting System are meant to help guide programmatic decisions. These scores should not be used to change the way an individual student is receiving instruction. WY-TOPP data should be used at the building and district levels to identify omissions or weaknesses in curriculum and instruction. Student information can be aggregated across the system to inform needed improvement efforts.

Remember: this is a single writing prompt on a single day. Student data should be analyzed in sets. A single score should not be the only measure used to inform needed support or enrichment for a student. Individual student information from WY-TOPP might be used in conjunction with classroom grades and district assessment data to determine what is best for that student.

Where can I find resources to help prepare my students?

WDE provides blueprints, rubrics, and exemplars for the writing prompts on the WDE website. Linked below are some of the resources teachers might find helpful in preparing for testing:

- [Writing Exemplars](#).
- [WY-TOPP Writing Blueprints and Rubrics](#).

Where can I find more information about best practices?

To see the [WDE Assessment Best Practices](#) guidance, go to the [WDE Assessment webpage](#).