AI-Assisted Responses to Statewide Assessment Writing Prompts

Vince Verges

Assistant Deputy Commissioner
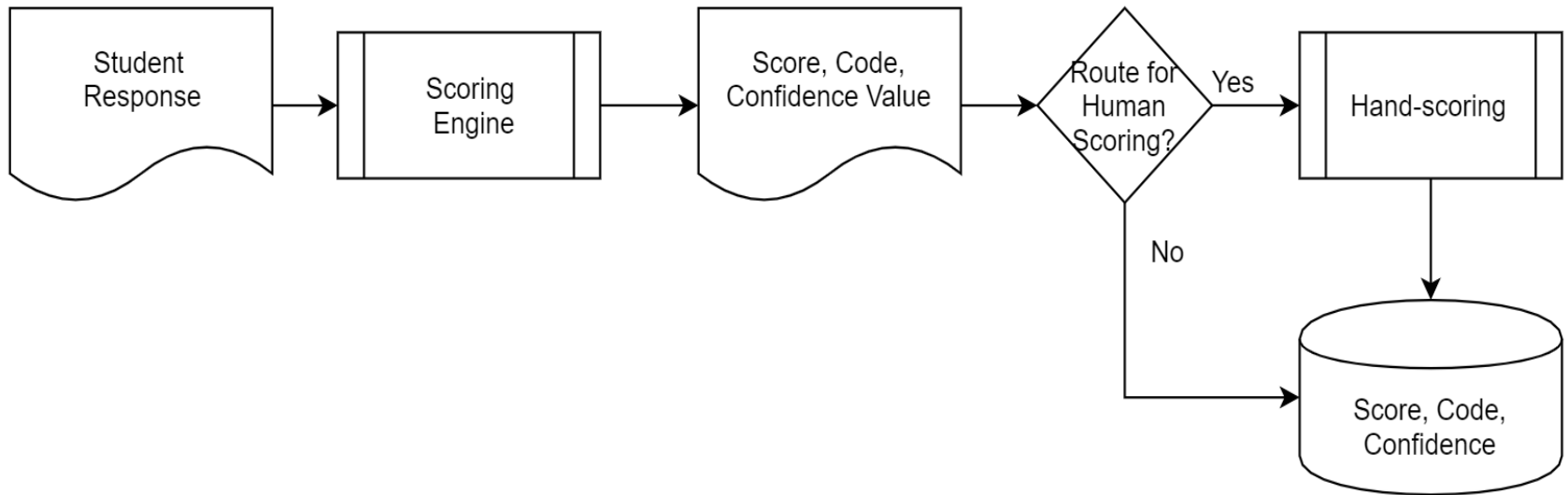
Division of Accountability, Research, and Measurement

FLORIDA DEPARTMENT OF EDUCATION
fldoe.org

B.E.S.T. Writing Scoring

Hybrid Scoring

Routing Flow

Responses routed for human scoring are scored by trained professional scorers.

AI-Assisted Writing Responses

B.E.S.T. Writing: AI-Assisted Responses

- Beginning in spring 2024, FDOE, Data Recognition Corporation (DRC) expert scorers, and Cambium automated scoring experts noticed an unusual pattern developing among certain responses to FSA Writing Retake and B.E.S.T. Writing prompts.

- The term Test Irregularity (TI) is used to describe a response submitted for scoring that contains unusual information.

- Human scorers at DRC are trained in the process of identifying these responses, and automated scoring processes may also identify responses that are then flagged for review by human readers.

AI-Assisted Responses – TI Criteria

TI criteria that may indicate a student has used AI technology in writing their response:

- Quality of the writing – The writing quality is distinctly different in one area of the response compared to another, e.g., the first half of the response is riddled with errors, poor word choice, sentence structure, and lack of development, whereas the last half is suddenly error-free, illustrative, and high-quality writing.

- Outside source suspicion – The response appears to include copied material from outside source(s), rather than or in addition to source material provided in the test stimulus.

AI-Assisted Responses: TI Criteria (cont.)

Additional TI criteria that may indicate a student has used AI technology in writing their response:

- Structure (including an essay title, which is unusual in responses to statewide Writing assessments)

- Language not typical of writing for the student's grade level

- Detailed information of which students are unlikely to have prior knowledge

- Highly specific conceptual similarities to those produced by Test Development Center (TDC) ELA content specialists in replications of AI-generated responses

AI-Assisted Responses:
Spring 2024 Identification Process

- FDOE and contractor staff identified a number of responses containing evidence of potential outside source use where the source could not be found. The issue affected a small percentage of responses but warranted further review.

- TDC ELA content specialists and FDOE staff members began typing prompts into ChatGPT and found that the same characteristics shared between student essays are also shared with essays produced by ChatGPT in response to the assigned prompt. These similarities became clearly recognizable over the course of reviewing hundreds of such responses.

AI-Assisted Responses:
Spring 2024 Identification Process

- Cambium's AS engine, which is trained on thousands of authentic student responses to each prompt, also separately identified responses from Spring 2024 testing that were flagged for TI and forwarded to human expert scorers for further review.

- More information on this process is presented on the following slides.

AI-Assisted Responses: Role of Autoscoring

- The AS engine currently flags unusual responses with condition codes and low confidence scores.

- These responses are then routed to human scorers for review. Human scorers apply the standard process for TI identification, which may result in identifying essays composed with the assistance of AI-generated source material.

- In other words, the AS engine does not determine whether a response has been generated by an AI platform. Rather, it routes unusual responses to humans for review.

AI-Assisted Responses: After Confirmation of a TI

- After the AS engine flags an unusual response, it routes to a human scorer and then to a team leader, scoring director, and the Writing content lead at DRC.

- When the response is forwarded to the appropriate content specialist at FDOE, the specialist will either confirm or reject the irregular aspects of the response, and the response either continues as a TI or is returned to the scoring pool.

- Autoscoring is covered in more detail in Section 5 of the [Statewide Assessments Guide](#).

AI-Assisted Responses:
After Confirmation of a TI (cont.)

- If the response is confirmed as a TI, FDOE content staff send a copy of the response to the FDOE Test Administration team with an explanation.

- The Test Administration team communicates with districts regarding these irregularities and requires an investigation, per Rule 6A-10.042, Florida Administrative Code (F.A.C.), Test Administration and Security.

AI-Assisted Responses:
After Confirmation of a TI (cont.)

- Investigations are conducted by school or district personnel, and evidence may be submitted to support that the response is original student writing.

  - Evidence may include FAST ELA Reading scores, samples of the student's writing or content-related report card grades.

  - The school or district may also uncover evidence of a student referencing an outside source during testing (e.g., a student during an interview admits to using a cell phone during testing). In these cases, the test score will be invalidated.

AI-Assisted Responses: Future Solutions

- FDOE teams, in conjunction with DRC and Cambium, will create AI-generated essays prior to each Writing test administration window using popular, free consumer platforms (e.g., ChatGPT, Gemini [formerly Bard], Bing, Claude).

- Cambium will build into the AS system the capability to compare student responses against those generated by consumer platforms and use this throughout the testing window and after the testing window closes to identify TIs.

14

AI-Assisted Responses: Future Solutions (cont.)

- When the AS engine detects elements described previously, it will route those responses to expert human scorers for review.

- FDOE is also investigating additional 3rd-party software products that are trained to identify AI-assisted responses.
  - There are several such products currently commercially available, but the reliability of any such product has not yet been fully explored.

- FDOE will continue to provide and strengthen training for test administrators to help reduce instances where students use unauthorized electronic devices during testing, which has long been a prohibited practice.
  - TDS, the online testing platform, does not allow internet access to other sites, nor does it have a copy/paste function for Writing tests. Students using AI-assisted material would have thus had access to an unauthorized device or aid during testing.

15

AI-Assisted Responses

An Illustration

AI-Assisted Responses – An Illustration

- The final portion of the presentation on AI-assisted responses will provide an example of an actual student response flagged for TI.

- The response was written during the spring 2024 FSA ELA Writing Retake.

- The prompt in question asked students to make an argument as to whether hard work or natural ability is more important.

AI-Assisted Responses – An Illustration (cont.)

- While the student response does not cite sources other than those provided, this is one of the indicators that can contribute to a response being flagged as a TI.

- For this response, the use of provided sources still has a bearing on the AI similarities noted.

- For this prompt, the sources were:
  - Source 1: "How Far Can Talent Take Us?" by Rebecca Williams. Written for educational purposes.
  - Source 2: Excerpt from "College freshman at age 9, medical degree at 21" by Bonnie Miller Rubin, from the Chicago Tribune.
  - Source 3: Excerpt from "Rethinking Mozart" by Andrew Moravcsik, from Newsweek.

AI-Assisted Responses – An Illustration (cont.)

- The example response was sent to the district, who in the course of their investigation interviewed the student.

- The student admitted specifically to using ChatGPT, which was apparently unobserved by the test administrator/proctor.

- Annotations to the response that follows refer specifically to ChatGPT, since this is what the student admitted to using.

- Color-coded student responses appear in italics, with corresponding annotations from FDOE/TDC appearing as bullets after each response excerpt.

AI-Assisted Responses – An Illustration (cont.)

The debate surrounding the role of natural ability versus hard work in achieving excellence has captivated scholars and enthusiasts for generations. As we navigate the <mark>complexities of this discourse</mark>, we draw insights from three compelling sources: "How Far Can Age 9, Medical Degree At 21" and "Rethinking Mozart." these sources shed light on the <mark>dynamic interplay</mark> between <mark>inniate talent and relentless effort in the pursuit of greatness.</mark>

- The words 'dynamic interplay' and 'innate' are not found in the provided sources. Most ChatGPT response iterations to the prompt employ these phrases in conjunction with some variation of a 'relentless pursuit of greatness/excellence' in the opening or closing paragraph.

- Note spelling/grammar errors as a result of manual transcription of AI response from an unauthorized device into the test platform, which does not allow for copy and paste from sources outside of the platform. Also of note is that these misspellings may appear in writing that addresses complex concepts.

AI-Assisted Responses – An Illustration (cont.)

A gift or a Limitation? The article "How Far Can Talent Take Us" explores the notion of natural ability as a double-edged sword. While innate talent can provide a significant advantage in certain endeavors, it can also be a limiting factor if not complemented by hard work. All of paragraph five in the same source is a great example of this. The story of a child may dazzle the world with exceptional abilities but the real test lies in wether these talents are nutured and cultivated through dedicated practice.

- These terms and phrases match those found in replicated ChatGPT responses.

- Nearly all iterations of reviewed ChatGPT responses to the same prompt highlight innate talent as a foundation or starting point, but that must be complemented by hard work. .

- Note spelling/grammar errors as a result of manual transcription of AI response from an unauthorized device into the test platform, which does not allow for copy and paste from sources outside of the platform. Also of note is that these misspellings may appear in writing that is otherwise quite complex.

21

AI-Assisted Responses – An Illustration (cont.)

The backbone of achievement in "College Freshman At, Medical Degree At 21" we encounter the remarkable journey of a young boy who defied conventional expectations. Despite possessing extraordinary intelligence and apitude, Yano's achievements were fueled by unwavering dedication and tireless effort. This narrative underscores the indispensable role of hard work in translating potential into tangible success.

- These terms and phrases match those found in replicated ChatGPT responses, especially so for 'tangible success'.

- Note spelling/grammar errors as a result of manual transcription of AI response from an unauthorized device into the test platform, which does not allow for copy and paste from sources outside of the platform. Also of note is that these misspellings may appear in writing that is otherwise quite complex.

AI-Assisted Responses – An Illustration (cont.)

*Unraveling the myth of the overnight success ==**"Rethinking Mozart"** **challenges the romanticized narrative of effortless genius**== by revealing the composer's path to mastery. Mozart's unparalleled musical genius was not solely a ==product== ==to== ==innate talent but also the result of rigorous training,== ==persistant== ==practice and pursuit of perfection. This== ==reevalutaion== ==serves as a potent reminder that excellence is not bestowed upon individuals but earned through== ==persistant== ==labor and excellence.==*

- ==The provided source mentions Mozart and a 'romantic' notion that geniuses are societally maligned. However, the romanticism mentioned in the source is related to societal misunderstanding rather than the effortlessness of talent itself.==

- ==These terms and phrases match those found in replicated ChatGPT responses.==

- ==Note spelling/grammar errors as a result of manual transcription of AI response from an unauthorized device into the test platform, which does not allow for copy and paste from sources outside of the platform. Also of note is that these misspellings may appear in writing that is otherwise quite complex.==

23

AI-Assisted Responses – An Illustration (cont.)

A path to excellence in the intersection of natural ability and hard work lies the blueprint of achieving excellence.

- The metaphor of travel from a starting point to the achievement of excellence/perfection serves as a common element of Chat GPT responses, particularly 'a path' that depends upon the 'intersection of natural ability and hard work.'

- At first glance, the idea being conveyed in this sentence may make sense, but in transcription, a word seems to have been omitted or mis-used here. Perhaps the intended idea was "…hard work lies in the blueprint…", or "…hard work is the blueprint…" As noted previously, these misspellings/errors often appear in writing that is otherwise quite complex.

24

Contact Information

ELA Contact Information

- **Sally Donnelly—ELA Secondary Content Coordinator**
  - Sally.Donnelly@fldoe.org
- Catherine Taylor—ELA Grades 9 and 10 Content Specialist
  - Catherine.Taylor@fldoe.org
- Gretchen Sims—ELA Grades 7 and 8 Content Specialist
  - Gretchen.Sims@fldoe.org

ELA Contact Information

- **Shakia Johnson, Ed.D—ELA Elementary Content Coordinator**
  - Shakia.Johnson@fldoe.org
- Ikema Morris, Ph.D—Grades 5 and 6 Content Specialist
  - Ikema.Morris@fldoe.org
- Stephanie Perry, Grades 3 and 4 Content Specialist
  - Stephanie.Perry@fldoe.org
- Angellena Jackson—ELA VPK–2 Content Specialist
  - Angellena.Jackson@fldoe.org

Other Questions?

Vince.Verges@fldoe.org

Susan.Lee@fldoe.org

Assessment@fldoe.org

www.FLDOE.org