

# How to Teach With AI

---

 [tomdaccord.com/how-to-teach-with-ai-grading-student-work](https://tomdaccord.com/how-to-teach-with-ai-grading-student-work)

## How to Teach With AI

---

### A Teacher's Guide to Grading Student Work With AI

---

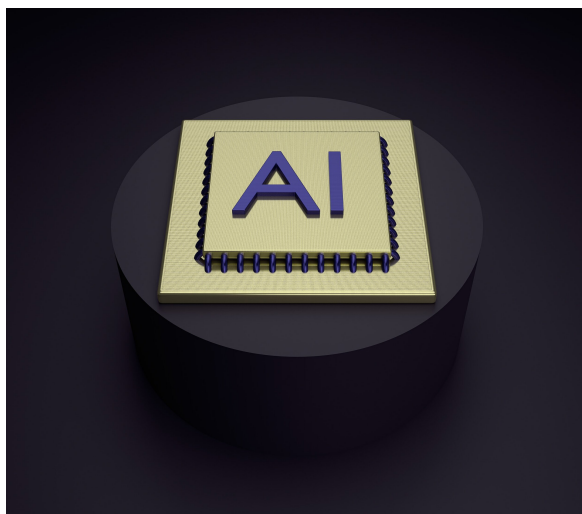
#### Understanding Grading with AI

---

AI for grading can be extremely helpful for teachers and save much time. For one, artificial intelligence algorithms are often able to analyze and evaluate student work efficiently and consistently. They can also provide immediate feedback and grade large amounts of student work quickly.

But grading student work is complicated and teachers use a variety of methods to assess student work. As a result, there are many variables for AI to consider when grading, which can lead to confusion or errors.

Here are a few common grading methods and how AI can help teachers grade student work:



#### **Unambiguous multiple-choice/fill-in-blank/true-false questions**

AI can grade unambiguous multiple-choice/fill-in-blank/true-false questions quickly and efficiently. That's because the "correct"/accepted answer to these types of questions is clear, singular and unwavering. There is a very small chance of a statistical error when AI is analyzing responses to these types of questions.

#### **Very short-answer questions** (one or a few words response)

AI can also grade very short answer questions quickly and efficiently — with some caveats. Very short answers typically constitute one or a few words. These do *not* involve sentence-length answers, let alone paragraph-length answers.

There can be some ambiguity in very short-answer responses that can be vexing. Let's say, for instance, that the "correct" response to a very short answer question is "President Theodore Roosevelt." But what if the student types "President Roosevelt"? Should that be accepted as a correct answer? There are two President Roosevelts in United States history, Theodore and Franklin. So, on the surface, an answer of "President Roosevelt" is wrong — or at least incomplete. But what if the students had not yet studied the latter president, Franklin, in their course? Is the answer "President Roosevelt" then acceptable to the teacher in this context? And what if students answered "Teddy Roosevelt"? Is that an acceptable answer? As you can see, there can many variables to consider when grading a very short answer question and AI may not know how to grade student answers without some guidance (more on that later).

### **Short-answer questions** (sentence to a few paragraphs response)

As you can imagine, there are more variables to consider when a student response is longer than a single word or two. Grading a short answer response of a few paragraphs appropriately can be a challenge for AI if it lacks relevant context.

That said, AI is excellent at analyzing written work for grammar, vocabulary, and syntax. It can draw from a huge reservoir of written academic work — from various sources and various grade levels — to analyze the structure, clarity, and effectiveness of student prose. That, in and of itself, is a huge benefit to using AI for grading student short-answer responses.

If the teacher is careful to provide the AI with relevant and appropriate context for the student answer then it can be a great help in grading short-answer responses.

### **The Five-Paragraph Essay**

AI is very efficient at analyzing the standard five-paragraph essay. That is because AI is typically programmed to recognize and evaluate specific patterns and criteria, such as those found in the five-paragraph essay.

According to experts from Georgia State University and Vanderbilt University, AI models "can identify and evaluate the lead, position statement, supporting claims and evidence in a five-paragraph essay "as well as a human." AI is well suited to evaluating "logic and persuasion," including how well a five-paragraph essay is organized and how well the arguments are developed.

So, it makes sense to leverage AI for grading the five-paragraph essay. But, as always, you should provide it with relevant and appropriate context to understand your expectations.

### **Extended Student Writing**

As mentioned, AI is programmed to recognize and evaluate specific patterns and criteria. For that reason, the more structured the writing, the more effective AI will be at analyzing it and grading it. So, extended writing that follows a common and recognizable essay pattern will be relatively easy for AI to analyze and grade.

For instance, many secondary-level essays begin with an introductory paragraph that includes a thesis statement, transition to supporting paragraphs with a topic sentence and supporting evidence, and end with a concluding paragraph that underscores the importance of the argument put forward in the essay. AI is very adept at analyzing and grading the organization, coherence, grammar, vocabulary, and syntax of such an essay as well as its logic and persuasiveness.

But AI struggles to grade written work that is out of the ordinary or creative, in some unexpected way. When essay organization and content do not follow predictable patterns AI can struggle to understand them. This is especially true of humorous or sarcastic content, which AI may not recognize as such. AI may also struggle with very nuanced writing. And it may struggle to grade creativity, for instance in a student poem. So, human oversight is essential when using AI to grade creative writing.

- **Numerical Answers and Equations**

AI is very strong when it comes to numerical calculations and solving equations. As such, AI can be of great assistance when it comes to grading answers that require numerical resolutions, as found in math assignments, some science classes, and statistics-based courses.

AI is also adept at problem-solving and can break down problems — numerical and otherwise — into a series of steps or building blocks. As such, AI is very useful for evaluating the problem-solving process that a student has undertaken.

Understanding and solving *complex* math has been challenging for AI, but AI can increasingly master math content with human-like reasoning skills. Recently, Google's Alpha Geometry demonstrated its ability to solve geometry problems at the level of an International Mathematical Olympiad gold medalist. And the Organization for Economic Co-operation and Development (OECD) expects that AI will soon master its international assessment of adults' numeracy and problem solving skills. So, AI will become even more useful for grading math responses as it masters even mathematical content.

### **AI Chatbot or AI Grader?**

There are two fundamental choices when it comes to grading student work with AI. One is to opt for a general AI chatbot like ChatGPT, Microsoft Copilot (Bing Chat), or Google Bard. The other is to use a dedicated AI grading system, as found at Gradescope, Quizgecko, TeacherMate, MagicSchool, Eduaide, and others.

- **Advantages of an AI Chatbot:** If you're comfortable creating AI prompts, and working with PDF documents, you should strongly consider using an AI chatbot for your grading tasks. For one, AI chatbots are more flexible than AI grading systems. You have more freedom in directing the AI what to do and in revising your instructions. Furthermore, you can upload important contextual documents — such as state standards, writing guides, rubrics, and more — that you may not be able to upload to an AI grader. (In theory, you can upload multiple documents to ChatGPT and other AI chatbots, but it doesn't always work. So, it's best to combine separate documents — ex. state standards document, writing guide, your rubric — into a single PDF for upload.) Third, the AI grader you are using may not be using an advanced large language model. Many AI tools hide the fact that they are running off GPT version 3 and not the more advanced GPT version 4.
  - **ChatGPT, Microsoft Copilot (Bing Chat), or Google Bard?** Which AI Chatbot you choose might depend on many factors, including timing. OpenAI's ChatGPT emerged first in 2022, and received a ton of exposure, so many early 2023 AI chatbot users were ChatGPT users. Furthermore, OpenAI introduced a more powerful version of its GPT language learning model (GPT-4) in March of 2023. But it began charging \$20 month for the use of GPT-4 and Microsoft's Bing Chat and Google's Bard arrived as free alternatives to OpenAI. Bard's large language model is arguably as good or better than GPT-4, but Google has been slow to release new chatbot features to the educational sector, relative to OpenAI and Microsoft. Microsoft's Bing Chat recently rebranded itself as Microsoft Copilot and features GPT-4, free of charge. Personally, I am increasingly using Microsoft Copilot over ChatGPT. That's not only because Copilot is free, but also because I have more immediate control over its response style and can create images directly with the chatbot.
  - For help crafting effective AI prompts, see my [Teacher's Guide to AI Prompts](#).

- **Advantages of an AI grading system:** The biggest advantage of a dedicated AI grading system is that it helps streamline the AI prompting process to make it more efficient. AI grading systems save a lot of time because they provide an interface geared directly towards grading student work and typically have access to a huge volume of grading resources. Instead of looking at a blank prompt box, as you would in ChatGPT or Copilot, AI grading systems feature specific and concise requests for educational information that help the AI system complete the grading task you desire as quickly as possible. Many of these AI systems also offer various assessment tools, such as quiz and test diagnostic assessment, essay grading and feedback, programming assessment, bubble-sheet grading, rubric creation, and standards-based assessment

**Quizzes and Test System v Essay Grading System.** AI Grading systems fall into two broad categories: Quiz & Test generators/graders and Essay graders. The former focuses on generating and grading quizzes and tests that include multiple-choice, true-false, fill-in-blank and very short-answer questions. The latter group focuses on grading extended written work and providing feedback on that work. There are exceptions of course, such as full-featured assessment systems found in (expensive) turnkey platforms sold to school districts and universities. There are also AI “teacher assistant” systems that provide a range of assessment options. But, for the most part, it’s not unusual to use one AI system to grade student tests and quizzes and another to grade essays.

In sum, if you are comfortable creating AI prompts and want more control and flexibility when it comes directing your AI grading, then I’d recommend you use an AI chatbot like ChatGPT, Microsoft Copilot, or Google Bard. If you’re willing to sacrifice a degree of control and flexibility for easier usage and faster speed, then my recommendation is to opt for a dedicated AI grader tool.

-Tom Daccord