IntelliMetric®
Questions and Answers

Table of Contents

✦ How does IntelliMetric work?

IntelliMetric is an essay scoring system. It draws upon sophisticated artificial intelligence, natural language processing, computational linguistics, and machine learning technologies to evaluate open-ended essay responses. It is the most sophisticated technology of its kind in the world, and you can use it in MY Access!®.

Think of IntelliMetric as a package of tools which evaluate more than four-hundred features of written language, from the simplest meanings of individual words to the most complex connections among sophisticated concepts. Six different electronic "judges" contemplate these features from different perspectives. Finally, the best decision based upon the different interpretations is rendered.

These tools do not evaluate writing in a vacuum. Instead, IntelliMetric is trained to evaluate essays written to specific writing tasks. During this initial training process, hundreds of essays are scored by human experts against a writing rubric. These essays are then presented to IntelliMetric. As IntelliMetric evaluates each scored essay, it learns how the writing features are associated with the score. For example, IntelliMetric may "read" one-hundred essays given a score of three. As it evaluates the features of each essay, it learns how to identify "three-ness."

This training process is repeated with essays from each score point until IntelliMetric can evaluate un-scored essays with human-level accuracy. The end result is commonly called a "scoring model." It is a "pool" of knowledge, unique to a specific writing prompt, grade-range, and genre, which IntelliMetric can apply to score new essays.

While artificial intelligence technology sounds complex, it essentially mimics the process human experts use to score essays: they study scored samples, learn to apply the rubric, and apply scores consistently over time.

✦ Is IntelliMetric scoring the same as grading?

IntelliMetric is designed to do one thing very, very well: it evaluates writing against a published rubric based on training it received from human experts. It will provide immediate, reliable scores each time a student clicks the "submit" button. This is its job.

Teachers, on the other hand, do many things well. One of these is to determine how students will use MY Access! and to communicate to both students and their parents how IntelliMetric will be used in the classroom.

Teachers decide, for example, whether a writing task is to be considered a "practice assignment," in which case IntelliMetric scores are visible in a student's portfolio but are not included in any reports of the student's performance. This means that the scores do not become part of a student's so-called "permanent record." Alternately, teachers may decide to hide IntelliMetric scores from the student's view altogether but retain them in teacher-level reports of the student's progress.

Most importantly, teachers decide whether to allow an IntelliMetric score to have a bearing on a student's final grade for an assignment. Since teachers can evaluate student writing or participation in class in ways that IntelliMetric never could, the process we call "grading" remains the responsibility of the teacher. Understood more broadly, it is the teacher's job to decide how MY Access!, and the IntelliMetric scores it displays, will be used in the classroom.

✦ Why is the holistic score not an average of the trait scores?

IntelliMetric provides a holistic score as well as five trait-specific scores in the areas of focus, content development, organization, language use, and mechanics. The holistic score represents the "overall" quality of the essay while the trait scores represent the qualities of more specific features of the writing.

When human experts score an essay using a writing rubric, they generally do not represent the holistic score as the average of the trait scores. Since IntelliMetric is designed to replicate the human scoring process, it works the same way.

It would not be wrong, per se, to calculate, report, and use the average of the trait scores. Certainly, the overall quality of an essay is related to the quality of its specific features. But human experts typically consider an essay's overall quality separately from the quality of its specific features. Why is this? Human experts recognize that individual writing features interact with one another in complex ways and can produce an overall impact that is greater or less than the average of their parts.

For example, consider an essay which displays a merely average control of mechanical rules, organization, and language style but argues its thesis genuinely and passionately. Because the student was writing under timed conditions, a human expert might score the individual traits low, but then apply an above-average holistic score to recognize the overall effectiveness of the writing.

Trained as it is to replicate human experts, IntelliMetric will apply scores in a similar fashion.

✦ How does IntelliMetric assess for voice?

The concept of "voice" is hard to precisely define. Typically, the term is used to refer to the qualities of writing that make it unique or distinctive. In other words, it refers to the ways the writer's personality is expressed through written words. However, if you were to ask twelve different teachers how they identify voice in writing, you'd be likely to hear thirteen responses.

So, can human experts be trained to identify and evaluate the qualities that determine voice? Probably. These traits are located within each of the writing traits evaluated by IntelliMetric, including focus, content development, organization, language use, and mechanics. Within each writing trait, the unique decisions a writer makes are what determines his or her voice.

To overcome the ambiguity about this concept, the writing rubrics in MY Access! settle the concept of voice within the "language use" trait. Thus, when human experts score essays to train IntelliMetric, they consider how effectively the writer uses language to convey his or her distinctive voice. So it is within this trait that IntelliMetric will identify and evaluate the writer's voice. Keep in mind, however, that an essay with an especially effective voice is also likely to be scored higher in other writing traits as well.

✦ How accurate is IntelliMetric?

As the gold-standard of machine scoring, IntelliMetric is at least as accurate as human experts, and often slightly better. To find out how this can be, let's consider how one measures the accuracy of essay scores.

When two human experts score essays, they are closely trained to apply the writing rubric. Their accuracy is measured by how often they apply the exact same score, adjacent scores, or scores discrepant by more than one point. Two well-trained humans are generally exact 65% of the time and adjacent up to 95% of the time, disagreeing by more than one point the rest of the time.

In numerous studies comparing the scores of IntelliMetric and a human expert, the rates of exact agreement frequently surpass 70% and the rates of adjacent agreement approach 100%.

What accounts for this? First, human experts are laden with bias—they are biased by the subject matter and by the relative value they assign to each of the writing traits. Some scorers, for example, punish spelling errors, while others do not. Accuracy is also negatively impacted by such factors as fatigue and the various interruptions that temporarily consume the scorer's attention.

IntelliMetric, on the other hand, is never distracted and doesn't harbor any pre-formed opinions. As long as its initial training is of high quality, it can score endlessly, applying the same rigorous analytics to each new essay.

Still, this is not to imply that IntelliMetric is more "intelligent" than human experts. Indeed, humans remain uniquely qualified to interpret and assess writing. What IntelliMetric does bring to the essay scoring process is an efficiency and objectivity that results in increased accuracy over time.

✦ Do MY Editor errors reduce the IntelliMetric score?

It is important to observe that MY Editor and IntelliMetric are separate tools which do not communicate with one another. When a student composes an essay on the student workpage and engages MY Editor, it evaluates the student's writing and suggests corrections for grammar, spelling, and usage errors. But because the essay has not been submitted for scoring, IntelliMetric doesn't know it exists yet.

After the essay is submitted for scoring, IntelliMetric analyzes the writing by referring to the training it initially received from human experts. In the course of its evaluation, IntelliMetric does identify errors in grammar, spelling, and usage and factors them into its evaluation of the essay. However, we must consider how IntelliMetric assigns "meaning" to the errors it finds.

The writing rubric avoids referring to a specific number of grammar, spelling, or usage errors at any given score point, because different error patterns can have a different impact on the overall writing. Instead, the rubric encourages human experts to determine how errors impact the communication of meaning in an essay. For example, numerous minor errors may have little impact on the meaning of an essay while a smaller number of very serious errors may significantly impact its meaning.

Thus, IntelliMetric will consider the impact of grammar, spelling, and usage errors and score an essay as a human expert would. This helps us understand why an IntelliMetric score does not always change after a student fixes mistakes identified by MY Editor— the score may go up or down depending on the impact of the errors. This also explains why an essay given the highest score may contain noticeable errors. While many people are biased against the number of errors, IntelliMetric remains true to its training and seeks out their impact.

✦ How can IntelliMetric score the content of an essay?

One of the writing traits that IntelliMetric evaluates is "content development." The MY Access! rubric defines the content development trait as, "The extent to which the response develops ideas fully and artfully, using extensive, specific, accurate, and relevant details."

To evaluate an essay's content, IntelliMetric draws upon the training it initially received from human experts. The scored essays used in this training process effectively teach IntelliMetric what concepts are in play, how they may be developed, and how they relate to one another. From this foundation, IntelliMetric can also evaluate "new concepts" that may not have been part of its initial training by drawing upon its own 16-million-word concept net. This neural network enables IntelliMetric to infer how new ideas relate to old ones.

As indicated in the rubric description, when human experts score the essays used to train IntelliMetric, they do consider the accuracy of the content. But they are not fact-checkers and cannot guarantee the accuracy of the content they evaluate. As their determinations are transmitted to IntelliMetric during the training process, it too is limited in its ability to fact-check the accuracy of content. Similarly, IntelliMetric is not designed to identify plagiarism, the deliberate copying of others' content.

For example, IntelliMetric will not recognize an error has been committed when a student writes that President Obama was elected in 1898. Rather, it considers how well the facts, true or otherwise, are used to support the thesis. For this reason, teachers are encouraged to do their own fact-checking when an assignment calls for it.

✦ Can I make IntelliMetric score "harder"?

Objectivity is one the greatest benefits of IntelliMetric scoring. Every time IntelliMetric scores a new essay, it refers to the training it initially received from human experts. The essays used to train IntelliMetric, in turn, are always scored against a writing rubric.

Consider how a teacher scores a student's response to a math problem. As a student works from one step to the next, a teacher can recognize precisely what the student has done well. But writing is a more subjective form of expression; there are an infinite number of ways one can tackle a "writing problem." How is a teacher to treat each student's approach?

A writing rubric lists performance characteristics at each score point for each writing trait. When human scoring is accurate, essay scores will be true to the rubric descriptions. A student can consult the rubric to objectively understand why he or she has received a certain score.

Thus, there is no way to change how "hard" IntelliMetric scores. Within MY Access!, teachers can change various feedback and readability settings, but they cannot make IntelliMetric score harder or easier—it remains true to the rubric embedded in its training at all times.

When grading a student, a teacher can easily "weight" the value of an IntelliMetric score by making it part of a larger assessment of a student's writing. In this way, teachers can make an assignment "easier" or "harder" by adding their own subjective evaluations to the objective IntelliMetric score.

✦ What does the "off topic" message mean?

The "off topic" message is one of a series of messages IntelliMetric may send to a student if it decides not to return an essay score right away. Called "non-scoreables," these reasons include: copy of the prompt, too short, insufficient, too many unknown words, repetitious, and major syntax problems.

Each time IntelliMetric evaluates an essay, it considers whether any of the features of the writing might impede its ability to apply a reliable score. Any unusual essay is flagged with a non-scoreable tag and then sent to a human expert for review. Within twenty-four hours of final submission, the human expert will decide whether the essay can be scored. These essays may also be used to further train IntelliMetric.

The "insufficient" and "too short" tags indicate that an essay might not be developed to the point where it can be accurately evaluated against the writing rubric. The "too many unknown words" and "major syntax" tags indicate that there might be problems with the language in the essay. The "repetitious" tag indicates that certain phrases might have been repeated, while the "copy of the prompt" tag is applied to essays that mimic too closely the actual language of the writing assignment. Last but not least, the "off topic" tag indicates that the essay may have strayed too far from the topic of the writing task.

It is important to note that a non-scoreable tag does not necessarily mean that an essay has been poorly written or cannot be scored. Rather, it simply means that IntelliMetric required an additional layer of expert human review to validate its score. Students are encouraged not to wait for the score, but rather to continue editing and improving the essay.

✦ Can IntelliMetric be tricked?

IntelliMetric is designed to do one thing very, very well: it evaluates writing against a published rubric based on training it received from human experts. It provides immediate, reliable scoring each time a student clicks the "submit" button. This is its job.

But, like its human counterparts, it is not perfect. Of the millions of essays scored annually, a small subset is scored higher or lower than its "true score." Generally, the essays that present the biggest scoring challenge are authentic submissions with unusual writing features. IntelliMetric does incorporate several tools to identify these essays; this is, after all, the purpose of the non-scoreable system which forwards unusual essays to a human team for review.

A second scoring challenge IntelliMetric encounters is a bit trickier. Clever writers frequently try to "trick the system" by submitting contrived essays for scoring. While these essays can take many different forms, a common tactic is to mix illegitimate text with legitimate text. IntelliMetric can successfully identify many of these "cheaters," but it cannot catch them all. IntelliMetric is designed to analyze writing and assumes the essays being submitted are legitimate. It wasn't designed to be a "truth detector."

Teachers should be aware of this limitation. Our human scoring team will review essays that fall into the first category within twenty-four hours. Essays that fall into the second category are usually harmless, but it is nonetheless important that teachers clearly communicate how their policy on academic honesty in the classroom applies to the use of MY Access!. Should teachers disagree with an IntelliMetric score, they can override the score and assign their own score to any student essay.

✦ What is the Approximator?

Approximator is a secondary scoring tool available in MY Access!. It is used to score writing prompts for which IntelliMetric has not yet been trained, including both pilot prompts and teacher-created MY Prompts.

The IntelliMetric training process is a highly focused one: the initial training essays scored by human experts are always specific to a writing task. Prompt-specific training provides IntelliMetric with a superior understanding of the "conceptual terrain" of the writing task, including what concepts are involved and how they are related. This understanding not only enables it to determine whether a response is "on or off topic," it is also crucial to evaluating the quality of the content developed in the essay.

Approximator, by contrast, is a very general scoring solution. It is not trained by human experts to specifically evaluate individual writing prompts. Instead, it is trained to evaluate writing across very general grade ranges, such as the elementary, middle, and high school levels.

The primary benefit of this scoring tool is that it can return reasonably accurate scores for custom writing assignments and pilot prompts. Students can write to virtually any task a teacher assigns and receive a holistic score. But this also illustrates the primary drawback of Approximator. Because it has not received prompt-specific, expert training, its scores can only ever be "reasonably accurate." Moreover, Approximator will never know if a response is off-topic. For these reasons, teachers are encouraged to be careful in how they use Approximator scores.

✦ What is ViAlert?

Any time IntelliMetric is asked to score a new essay, it begins by looking back at the training it initially received from human experts. In addition to evaluating the features of writing that will enable it to apply a reliable score, it also considers whether the essay contains language indicating planned violence by the writer against him or herself or another person.

This "violence alert" feature is called ViAlert. Like the other tools used by IntelliMetric, this one is "smart." It does not simply look for certain catchwords or phrases, such as "kill" or "gun." After all, if it were this simplistic, it would flag every essay written about Shakespeare! We all know that students have good reasons to write about violence, so ViAlert looks at the context of the writing and identifies language within that context that is suggestive of inappropriate violence.

There are no scoring "penalties" assessed by ViAlert. IntelliMetric evaluates and reports scores for essays flagged by ViAlert to students, but it will also highlight these essays in the teacher's portfolio with a yellow warning indicator. Because they are in the best position to know their students, teachers are encouraged to review these essays carefully to determine the writer's intent. If further action is required, the teacher can best decide how to proceed.