turnitin

WHITE PAPER

Turnitin Revision Assistant: Behind the Scenes

Elijah Mayfield and David Adamson

Table of Contents

Revolutionizing the Experience of Writing to Learn

Foreword

Turnitin Revision Assistant is a groundbreaking classroom product for teaching writing. We are very excited to be releasing it to the public at large after two years of development and classroom pilots.

However, the technology behind this product would be familiar to computer scientists in any advanced research institution.

So let's start with a confession: from a technology standpoint, nothing that we've built here is particularly revolutionary. Revision Assistant is built from algorithms that have gestated in universities for years. There are no robotic brains here. We have not discovered a new form of artificial intelligence. We aren't even building a self-driving car.

Our innovation is in trusting teachers to tell us how to use these algorithms for the benefit of their students, and trusting that with the right support, students will eagerly take ownership of the writing process and strengthen their writing skills. For the first time, natural language processing algorithms are being put into an application built by educators. It's the first time that a product built on machine learning has been designed directly in the classroom.

The team that designed Turnitin Revision Assistant is made up of current and former teachers. In 2014, we developed an initial prototype and offered it to a handful of institutions. And then we did something that very few machine learning researchers would ever do: We went back to school. We sat in classrooms and watched what happened when teenage students got feedback on their drafts. We took copious notes and brought them back to our office. After months of research, we spent the next summer and fall rethinking many of our assumptions.

In spring 2015, we rolled out a new iteration of Revision Assistant, this time with a richer user interface, as well as a more refined system for providing writers with actionable feedback. And then we went back, again, to spend more time in the classroom observing students and talking with them. We asked for their harshest critiques. Luckily, eighth-grade students are not shy.

The product launching in January 2016 is thus the result of years of student and teacher feedback. This paper discusses our product development process in greater detail: the collaboration to understand what high-quality writing prompts look like for educators, the expert partnerships to create feedback that is meaningful to students, and the technology that makes this content available on demand.

We're thrilled to share it with you.
— The Authors

About the Authors

Elijah Mayfield has been working in computational linguistics and machine learning since 2008 and founded LightSide Labs to focus specifically on improving automated feedback for student writers. In 2014, the LightSide team joined Turnitin, where Elijah now serves as Vice President of New Technologies. He leads a Pittsburgh-based team of teachers, researchers, developers, and designers that are building better tools for teaching writing.

David Adamson taught for six years at Baltimore City's Digital Harbor High School before coming to Carnegie Mellon University as a Ph.D. student in Language Technologies. During his graduate studies, he studied effective ways for intelligent, automated tutoring systems to interact with students in positive, social ways that improved learning. David co-founded LightSide Labs and now leads the Turnitin New Technologies research team.

1.0 Executive Summary: What is Turnitin Revision Assistant, and How Does It Work?

1.1 Introducing Turnitin Revision Assistant

Turnitin Revision Assistant is an online tool that provides instant feedback to improve student writing throughout the writing process. Our goal is to support teachers in providing engaging, high-quality writing instruction and to motivate students to write more, revise more, and grow their writing skills.

Revision Assistant grounds the writing process in great prompts, which teachers can select from a library and assign to their students. When students receive an assignment, they can use Revision Assistant to plan and draft their work before calling for a "Signal Check." Signal Checks provide students with instant formative feedback that they can use to begin revising and improving their drafts right away.

Revision Assistant's feedback is multifaceted and unique. First, the Signal Check provides a student-friendly visualization of trait-level scores on standards-aligned rubrics. Second, actionable, direct comments appear in the margins of students' compositions and are tied to sentences highlighted in the text. This feedback emphasizes areas of strength, as well as offering strategic suggestions for sentence-level improvements. The combination of prioritized positive and constructive comments encourages students to continue revising and progressing. Students can continue to revise and get feedback through additional Signal Checks as many times as they would like before sharing their work with their teacher.

The first time educators and administrators see Turnitin Revision Assistant's instant feedback, their first reaction is often one of excitement as they imagine how it could support instruction and learning in their classrooms. Their second reaction is often one of curiosity: How does Revision Assistant deliver these comments? How does it know which portions of an essay to target for feedback? Who wrote the comments, and how does Revision Assistant match this feedback to individual students' drafts? This paper addresses these questions and delves into the development of Revision Assistant's content and technology. Some of the vocabulary in this chapter's summary might be unfamiliar on your first read; progressing through the document will help to put together the pieces.

1.2 How Turnitin Revision Assistant Works

Before a **prompt** ever appears in Revision Assistant's library, we partner with schools who share feedback on prompts that we've written internally, propose prompts of their own, and help us align our curriculum to their school's experience. Internal experts and classroom teachers collaborate to ensure that every prompt is clear, understandable, and interesting to students. Each prompt is associated with a **rubric** that is aligned to learning objectives for a specific genre of writing.

Then, we ask a small number of teachers to assign each proposed prompt to their students. This gives us a few hundred example essays that serves as our **training set.** This collection of essays shows us the range of writing that students will likely produce in response to that prompt and provides a foundation for how future essays will be evaluated.

After the training set is collected, we partner with assessment experts, who evaluate the essays against the prompt-specific rubric in a fair and reliable way. We then use these scores to create a computational representation of the essays that received them. By analyzing the **features,** or building blocks, contained in student drafts that received a particular rubric trait score, we can create a **scoring model** to accurately evaluate future student essays written about that same prompt.

Once this training process is completed, the prompt enters Revision Assistant's library and can be assigned as part of regular classroom instruction. Any teacher can select a prompt aligned with their curriculum, assign it to students, and have students submit their essays directly through Revision Assistant. These essays can immediately be **scored** by Revision Assistant—which is really a comparison with what the scoring model has seen before in the training sets.

At this stage, Turnitin's **feedback engine** enters the picture. This is the key element of Revision Assistant: Our algorithms allow the software to provide students with actionable feedback that helps them immediately chart a path forward. After using **validators** to check that an essay is on-topic and written in good faith, the feedback engine dissects an essay into its component parts. Each passage of the draft is evaluated for its impact on the overall text's quality. When an area is clearly stronger or weaker than its surroundings, the automated system recognizes this and **highlights** the passage for feedback. This highlighted text is then paired with an appropriate **comment** from a database that has been carefully created and catalogued by veteran educators. The content, tone, and specificity of this feedback is unique to Revision Assistant. Students can immediately put it into practice, revising and re-submitting their essays for additional feedback from Revision Assistant until they are confident they are ready to turn in their best work to their teacher.

How Turnitin Revision Assistant Delivers Specific Feedback

Our goal in all of this work is to ensure that Revision Assistant empowers students and teachers with fast feedback that is helpful and fair. This white paper explains each moving part in this process in greater detail, including the work we do with partners to ensure that our technology and curriculum align to support writing instruction like never before.

2.0 The Content and Curriculum of Turnitin Revision Assistant

Providing high-quality feedback on a student's writing requires understanding what students and educators believe is important and how they define success in context. These values intersect with many components of curriculum and assessment.

In this section, we describe how Revision Assistant is built on expertise from the classroom to define the most crucial pieces of that puzzle.

2.1 Prompts

Why is a prompt-based approach key to making instant feedback effective?

Revision Assistant is a curriculum product that provides teachers with writing prompts. By requiring that all students using Revision Assistant respond to a specific prompt, we can ensure that the feedback we provide is sensible and targeted to the appropriate mode of writing. From our perspective as developers of Revision Assistant, approaches that do not consider the writing prompt lead to limited feedback that is blind to mode, genre, and audience. In other words, ignoring the context of a prompt leaves us with generic feedback—for instance, the types of corrections on spelling, mechanics, and usage that a word processor provides today.

How are Revision Assistant's prompts developed?

To ensure that Revision Assistant can support writing instruction nationwide, Turnitin has partnered with recognized educational nonprofit organizations like WestEd and the College Board® to create an initial wave of writing prompts that fit in a modern curriculum for writing. Our district partners have helped us further modify and extend the prompts. Partner educators have suggested new writing topics and have helped position prompts from Revision Assistant's library within the kinds of units that teachers are encouraged to enact throughout the year. We will continue to expand the prompt library as more and more schools nationwide partner with Turnitin, with a goal of annual refreshes and up-to-date content for users of Revision Assistant.

Example Prompt

Informative Essay (11th–12th Grade)

Examine the targeted advertisements that you see when you use social media and consider whom advertisers seem to think you are. Compare this "advertising identity" with your own view of yourself. Write a post for a class website analyzing similarities and differences in your "advertising identity" and your "real identity," using examples from your social media activity and from your offline achievements and interests.

2.2 Rubrics

What role do rubrics play in Revision Assistant?

Students' responses to these prompts are assessed based on a rubric to ensure that all evaluation is rigorous, fair, and meaningful. We launched Revision Assistant with rubrics that incorporate best practices for various ages and skill levels in three genres: narrative, argumentative, and informative writing. Unlike long-form research papers and poetry, which cannot be evaluated automatically today, short texts aligned to these three genres can be addressed by our software's capabilities to provide specific, actionable feedback to students.

How are Revision Assistant's rubrics developed?

To design our rubrics, we examined each state's learning objectives and focused on aligning criteria to standards. We also analyzed our prompts to ensure that we selected the rubric that made the most sense for each writing task. After selecting criteria, we needed to define score level cut-offs that were reasonable, reliable, and fair. A frequent downside of many rubrics is that they fail to give actionable advice. For example, a 4-point rubric for language may designate a score of 2/4 for "somewhat vivid" and a 3/4 for "vivid." That doesn't necessarily help students who are trying to understand where they are and how to improve. To address this, we defined each Revision Assistant rubric's score levels in terms of positive, specific writing attributes whenever possible. These rubric definitions need to be backed up with data in order to use them as the basis for providing feedback. For this, we turn directly to student texts.

Example Rubric Criteria and Levels for an Argumentative Essay (6th Grade)

	Claim and Focus
Advanced	The writer clearly introduces an arguable and specific claim or analysis based on the topic or text(s) and remains focused on supporting that claim throughout the essay.
Proficient	The writer clearly introduces a claim or analysis based on the topic or text(s) and stays focused on the claim throughout the essay.
Developing	The writer's claim or analysis is discernible, but it may not be clearly stated in an introduction. The writer may sometimes lose focus on the argument, instead summarizing the topic or text(s).
Emergent	The writer does not clearly make a claim or offer an analysis, or the claim is overly general or vague. The essay does not maintain a focus on an argument or the topic or text(s).

2.3 Training Sets

What is a training set, and why does Revision Assistant use them?

The most important data for providing effective feedback on student writing is examples of real student work. We model that in computational terms with training sets. A training set is a collection of essays that are written by students from the target population, in response to a selected prompt.

When working with schools to collect a training set, it is important to include a broad range of responses that represent the vast majority of the ways that students might interpret and respond to the prompt. Typically, we try to collect at least 300-500 sample essays for a training set, across a narrow band of grade levels that are appropriate for each prompt. We gather these drafts across multiple classrooms—and ideally across multiple buildings—for diversity. With a smaller number of samples, a model can often fail to account for the more unusual choices that students may make in their writing. If we train a model with hundreds of essay samples, however, there is usually enough variation and evidence that the model can distinguish patterns in student work. After reviewing five hundred unique student responses to focused prompts, there will be very few future cases that bear no resemblance to previous texts.

How does Revision Assistant create its training sets of student writing?

For each prompt available at launch, our essay training sets came directly from real classrooms. We partnered with enthusiastic teachers at districts where we have a strong, ongoing relationship with the administration and the curriculum team. These partners got an early window into the prompts that would be available in Revision Assistant and assigned them to students. Students' responses were accumulated into training sets to showcase authentic classroom work.
In the case of prompts associated with AP® courses, we received secure access from the College Board to anonymized student essays from the AP® exams held at the end of each academic year. We take great care to ensure student privacy when building all our training sets: students' names and other personally identifiable information are not attached to the writing samples provided.

2.4 Scores

How is a training set of essays assessed to lay the groundwork for automated assessment and feedback?

Defining the right process for evaluating a new body of student work on these rubrics is the critical next step. Ensuring reliability when evaluating a large set of essays written to the same prompt is a challenging and important step to providing fair, meaningful feedback.

To get to reliable scoring without bias, we partner with assessment experts with a deep education background. For our initial set of writing prompts, our partner is the Center for Assessment at the University of Georgia, which has specific expertise in reliable and valid assessment development and implementation. These experts work directly with current and former teachers to define a scale for evaluating student writing that is aligned with each prompt's topic. They take into account both the original intent of the prompt and the skill level of essays we provide to them, under a confidentiality agreement that protects student privacy. We collaborate with the Center to review our rubrics and prompts, discuss the texts in our training sets, and build up a reliable set of scores for each essay along each rubric trait.

Why can Revision Assistant analyze essays, but not research papers or poems?

The process outlined above is built on several assumptions that are aligned with particular types of writing instruction. Specifically, Revision Assistant is designed to support writing tasks that different teachers are likely to evaluate consistently, as well as assignments for which students are working towards the same goals. Before moving on, it's worth examining both of these choices.

Consistent Feedback Across Instructors

As we built Revision Assistant, we used a clear definition of "reliability" in scoring an essay: Two different readers of the same text rate the essay the same way on a rubric. In the end, we make a philosophical claim: A rubric is ready for use with a prompt if two independent readers consistently give a student the same feedback. Our scoring model is ready to evaluate new student essays if it can replicate that same level of consistency.

This approach does not suit assignments that are fundamentally subjective. Assignments that do not expect the consistent application of a rubric by two or more people during reading are not a good fit for our approach. This includes poems or free-form writing exercises. In other words, it is appropriate to use Revision Assistant only for prompts for which the responding essays will be evaluated by clear, easily understood criteria.

Common Goals Across Students

Similarly, the collection of training sets implies that Revision Assistant will look for the same strengths in writing across all students. This includes the critical assumption that students will, broadly, bring the same types of information, background knowledge, and rhetorical style to an assignment. There are assignments, such as independent, primary research, for which students' writing styles are very likely to differ, or for which the subject matter will be unique to each paper. We generally avoid recommending Revision Assistant for these assignments.

2.5 Summary

The prompts, rubrics, training sets, and reliable scores we partner with teachers to produce serve as a foundation for Revision Assistant's automated feedback.

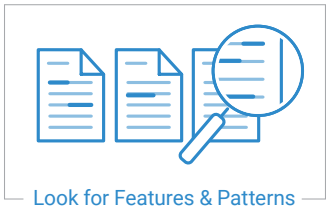
In the next section, we dive into how these sources of evidence are pulled apart and examined quantitatively to model and understand students' writing in context.

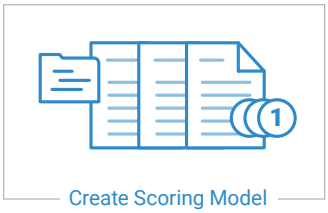
3.0 Approach to Analyzing Student Writing

In today's world, we expect computers to make very precise decisions in a very messy universe. For example, we expect our email programs to neatly sort spam into a trash folder without human intervention. We take for granted that services like Netflix or Amazon observe our online habits and suggest what to watch or buy next. In many aspects of daily life, technology-based recommendations have become ubiquitous.

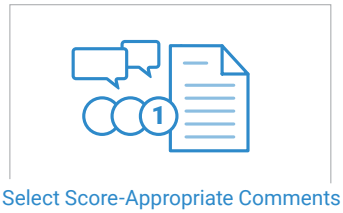
Automatically making clean decisions in a messy world is only possible through a set of algorithms collectively known as "machine learning." In this framework, the software algorithm looks at a long history of past actions (after being anonymized to preserve user privacy)—for example, a large set of emails or shopping sessions across many users. When a new piece of data comes in, the algorithm compares it to the most relevant examples it has seen in the past in order to make its next decision on how to perform. The resulting behavior looks highly intelligent, but is really just built on straightforward responses tied to matching patterns and mimicking the prior expertise of individuals. Below, we explain how Revision Assistant harnesses this kind of pattern recognition to analyze the quality of student writing and target specific areas for improvement.

Collect & Sort Training Set


Look for Features & Patterns



Create Scoring Model



Select Score-Appropriate Comments

3.1 Dissecting a Training Set

What data from essay training sets does Revision Assistant analyze?

Automated assessment is essentially sorting essays with a particular rubric criterion score into piles and analyzing what each pile has in common.

The definition of these piles comes from the educators who wrote the rubrics and scored the essay training sets, as described above. By the time a computer gets involved in predicting scores on new essays, its task has been simplified greatly: comb through all the characteristics of each essay in the training set and compile them into lists.

We call characteristics of a text *features*. When the software tabulates these lists, it pulls out *every* feature of an essay that it can find: every word, every syntactic structure, and every phrase. A typical student essay contains many thousands of features that can be used to inform its score and the feedback it receives.

Our approach looks for proxies that can stand in for a student's intention and meaning. Features that correlate well with training set scores can effectively predict future scores. If some aspect of a text is not included in this list of features considered by the algorithm, then by definition, that aspect of the text is not considered in the assessment of the essay. Getting the features right is crucial.

Our approach is inclusive, and we attempt to keep as much detail about a text as possible. This is a major difference from past approaches to automated assessment: we do not predetermine features that correlate well with standardized test scores, nor do we rely on reductive measures like word count or reading comprehension difficulty level of a text. Instead, we take a discovery-based approach, examining both individual features and combinations of multiple features that appear in a training set to find out what suggests success for a particular writing task.

For a detailed example of how the linguistic features embedded in a student's essays can shed light on their writing choices, please see Appendix A.

3.2 Looking for Patterns

Which facets of a student's writing impact the scores and feedback Revision Assistant provides?

None of the information pulled from a text intrinsically indicates its quality. Software cannot guess, on its own, which of the thousands of features are signs of a strong writer, and which should be praised. Luckily, with Revision Assistant, we don't need to pass judgment on which features are important. Instead, we trust that the patterns become meaningful in context, after observing the assessment of a wide range of authentic student texts in the training set.

We hypothesize that some writing characteristics—either content or style—are consistent across students at each skill level. If this is true, then some features will be very common among essays written by struggling students, and those same features will not appear in essays by the most skilled writers. The converse will also be true: features that overwhelmingly appear in the best essays are likely to be rare or nonexistent in drafts that need substantially more work.

The accumulated evidence from all of those features can be weighed, like a scale. A single essay will contain thousands of features. Some will be evidence that suggests a stronger essay, and others will be evidence of a low-scoring essay. The balance of all of this evidence will dictate the final rubric score.

Many features, though, will have little or no information. It tells us very little, for instance, when a writer uses words like “the” or “and,” so these features are given almost no weight or discarded entirely. The resulting set of weights for each feature is unique to each rubric trait and is derived from first principles for each new prompt and training set that we collect.

3.3 Creating Specific, Accurate Models

Why does Revision Assistant create a new scoring model for every prompt?

The qualities of good writing cannot be defined just once and applied to all future writing prompts. Each time we collect a new training set, the features that are present in student essays are different from any set we’ve collected before. Every time a training set is scored in response to a new prompt, on a new rubric, the features end up being realigned along that rubric.

Strong characteristics of one genre of writing may or may not transfer over to new domains. The only way we can determine which features matter most for a particular assignment is to repeat this process, relying on the representative sample of student texts that gives us an accurate lens into the classroom and the expertise of the educators that created the rubrics and scored the training set.

The final assignment of a score to a text tells only a tiny fraction of this story, though. By its nature, the process of assigning a number to a text is intensely reductive. By saving all of the information that we collect while deciding on that number, though, we give our software a treasure trove of information about student process and the work they perform. That data, if properly packaged, can then be put to work on the true goal of the product: giving appropriate feedback to empower student writers and help them improve their writing skills.

4.0 Approach to Providing Instant, Actionable Feedback

The methods in the previous section make sense when a student’s work fits into the progression of ability and skill levels that we’ve seen in our training data. By our definition of reliability, we ensure that each writing prompt has an appropriately paired set of features and weights, ready for automated assessment. If our interest was purely in scoring, we would call this success. This is what the education industry has historically accepted. Thanks to advances in technology and collaboration with classroom experts, we can do better.

By examining features to assign scores on a rubric, we have collected reliable, targeted sets of the hundreds of choices that a student made while crafting their essay. These texts are usually concise - a few hundred words or less. Now, the real question becomes how to motivate those students to improve based on what we’ve found.

Our solution again is to trust in teachers to guide the way for us. This section breaks down our approach to formative feedback, using the evidence from automated assessment as a starting point for promoting meaningful change.

4.1 Recognizing Red Flags Fast

How does Revision Assistant encourage students to write on-topic and stay on task?

Our scoring models have been trained on examples of authentic student writing, but that's not the only kind of input Revision Assistant will encounter in the classroom. Students might paste in source texts wholesale, ramble off topic to kill time, or enter gibberish into the system. Instead of trying to score these texts in good faith, we take a cue from teachers and try our best to bring students back on task.

Before assessing any essay, we first run it through a gauntlet of "validators," and only provide feedback on essays that make it all the way through. A student whose essay doesn't pass validation is not assessed; their writing is not ready for detailed feedback. Instead, Revision Assistant provides the student with a tailored message about the snag they've hit, with general guidance on moving toward a text that could be scored and commented on with confidence. Teachers also receive a message explaining the reason a draft did not receive feedback.

For more details on how Revision Assistant validates student texts and helps students stay on topic, please see Appendix B.

4.2 Targeting Sentences for Feedback

Why does Revision Assistant highlight specific parts of students' drafts?

Software that scores essays isn't terribly novel. Software that gives feedback to help students become better writers, on the other hand, is unique. We aim to do it right. This requires going beyond serving up a plethora of generic advice for students to muddle through and instead helping them strategically target specific portions of their work.

Within a student essay, Revision Assistant highlights sentences that show evidence of stronger or weaker writing, relative to the rest of that essay. These highlights both focus the student on areas where she can improve and celebrate sentences that can serve as exemplars for the student to emulate elsewhere.

How does Revision Assistant identify which parts of an essay are strong and which ones need the most attention?

Revision Assistant measures the impact of each sentence in an essay along every rubric trait. As described above, we begin by determining the essay's overall score on that trait. Then, we look back at the feature weights used to make that decision and determine which individual sentences most strongly influenced that trait score; we can do so because of our unique approach to scoring, which maintains our full set of features and their original locations in the text.

Any one sentence is unlikely to change the overall score that a text would receive, but each sentence contributes to the whole and has an impact. For providing feedback, we model this impact on a sentence by sentence level, by reconsidering what the text would look like if the sentence was absent. Would the score change? Even if the score would stay the same, would the text be weaker or stronger? By ranking all of the sentences in the student's essay based on how they uniquely contributed to the final assessment (either for better or worse), we can prioritize our highlights. This also gives us an opportunity to encourage concision in student writing; for new writers, removal is often harder than addition of new content. With this approach, Revision Assistant encourages students to reflect on where they can cut or rework their text, not just continue to add new ideas.

4.3 Prioritizing Areas of Focus

Which sentences does Revision Assistant highlight for feedback?

Now, with a ranked list of sentences on each trait, Revision Assistant must choose where to focus student attention. Our software provides targeted support to each student in their Zone of Proximal Development. In our ranked list, weak sentences are examples of writing that's holding a student back from reaching the next level. Strong sentences provide the student with reachable exemplars to guide their improvement throughout the rest of their writing.

Once we have a ranked-by-impact list of the relatively strong and weak sentences in an essay, we pick a small handful of the most impactful to focus our feedback on. For each round of feedback, we choose four sentences to highlight. Our research over the last two years has shown that more feedback is not always better feedback; rather, encouraging iteration through multiple revisions is more valuable than providing feedback throughout an entire text at once.

Instead of just picking the four highest-ranked sentences, though, Revision Assistant coordinates its selection of comments across traits, and always attempts to strike a balanced tone. Specifically, RA aims for a mix of feedback styles, based on a series of constraints. Instead of just picking the four highest-ranked sentences, though, Revision Assistant coordinates its selection of comments across traits and always attempts to strike a balanced tone. Specifically, Revision Assistant aims for a mix of feedback styles, based on a series of constraints.

Now, we've isolated the areas in a student essay to focus on. We need something to say—this is tricky. Computers don't generate original thoughts, and in fact, are not interpreting the student's text at all. Again, teachers have given us the help we need.

Constraints When Selecting Feedback

- Both strong and weak sentences are highlighted in each round of feedback, usually selecting two areas of improvement and two of the student's best sentences.
- Typically, each round of feedback will provide one piece of feedback on each rubric trait, unless there is a strong reason to provide multiple comments on a single trait. This might occur when a student is struggling on one particular trait or that trait interacts with the overall structure of their text, in the introduction or conclusion.
- Any one sentence never receives more than one comment in a single round of feedback. For example, a sentence won't simultaneously be marked as "Weak in Clarity" and "Strong in Use of Evidence", as contradictory feedback can confuse the students in need of the most practice. Multifaceted feedback on the same sentences can instead be received through multiple Signal Checks.
- If a draft is submitted multiple times with no change, or very little change, we provide the student with new feedback. Students should never get stuck just because they're hearing the same feedback again and again.
- Sentences at the start and end of a text are more likely to receive feedback because of their influence on an essay's introduction or conclusion; Revision Assistant's comments include specific advice about how to introduce the reader to a text and close an essay with a strong conclusion.

4.4 Providing Specific, Actionable Comments

How are Turnitin Revision Assistant's comments developed?

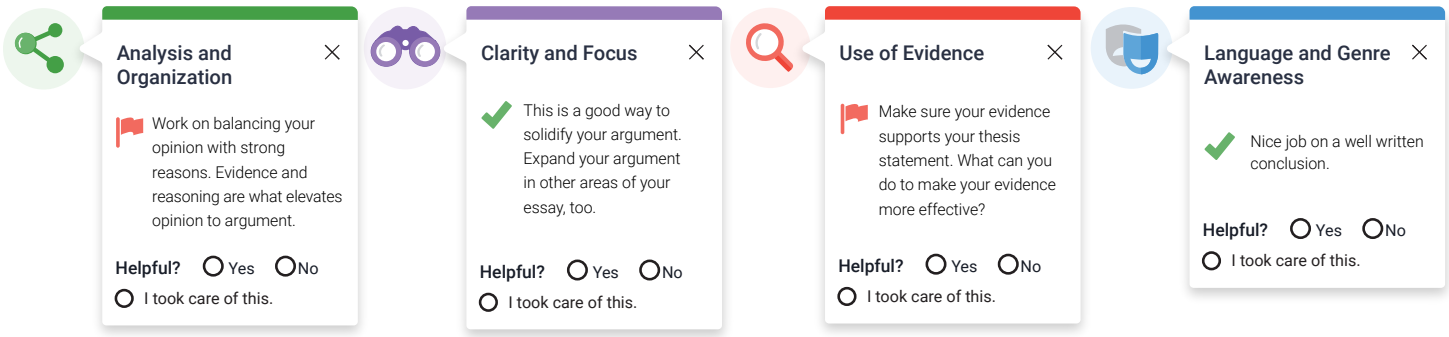
Revision Assistant's comments are *not* written by a computer, but rather created and curated by experts. All comments are written by our in-house content team comprised of veteran educators and composition experts. Currently, that team includes a veteran high-school teacher and literacy coach, a professor of literature and composition, and the former editor of a leading writing handbook. Developing comment templates for each rubric is a thorough, iterative process. When we have access to the rubric or prompt authors, we include them in our feedback loop.

Each comment is linked with a specific rubric trait for a specific genre, offering actionable strategies for reflection and revision matched to students' current level and the writing task at hand. Our content team also develops comments that apply to specific parts of a draft, such as the introduction, the conclusion, and text containing quotes from sources included with the prompt.

Comments have to be specific enough to be actionable; this means they must be authored to be applicable in new student texts that the content team has never seen before. This takes a long time and many iterations using anonymized texts in the training set for each prompt. Comments are tuned for tone and breadth to match the needs of the students represented in our training set. Once again, this underscores the importance of quality prompt and rubric writing, representative student texts, and reliable, consistent scoring of a training set to ensure meaningful assessment and feedback.

We've written over 1,000 comments in this collection, and more are created with each new prompt added to the curriculum library. This allows us to select from and surface the right comment for the right student at the right time during the writing process.

Example Comments



How does Revision Assistant match comments to specific sections of student writing?

For each sentence that has been flagged for feedback, Revision Assistant winnows down the pool of possible comments to identify those that are the best fit. For example, a given sentence may be a relatively "weak" sentence in an essay scoring a "2" on the "Development of Ideas" trait in our "9th-10th Grade Informative" rubric. In this specific scenario, only a handful of carefully targeted phrases would offer appropriate strategies for students in that niche.

Each essay written to a given rubric draws from the same pool of comments; comments developed for different genres of writing, by contrast, have fairly little overlap. Occasionally, comments are automatically tailored to include prompt-specific details, like the author or title of the source document included for an informative assignment. In other cases, comments will include reminders about the themes or strategies that a specific prompt asks the student to focus on. In general, the feedback that a student receives will be appropriate to all of their writing within the assigned genre, helping students develop effective strategies that will transfer to new writing in the future.

Multiple comments may address the same strategy, rephrasing the idea in different ways. This means that students can request multiple Signal Checks, re-reading their text and considering the feedback they receive until a comment's advice resonates with them. Students do not receive the same set of comments twice, so they never get stuck.

4.5 Instant Feedback Creates Meaningful Conversations

How does Revision Assistant shape the student-teacher dialogue about writing?

When students are asked to think critically about their own writing, we observe them asking teachers for help. The teacher becomes their guide and collaborator, as opposed to the wielder of the red pen. Students can work with greater independence throughout the drafting and revision process, turning to their teacher for deep workshopping and final feedback. This is not a failure of the product—in fact, it is an explicit goal of our work.

The teachers that we know put their full effort into helping their students. Revision Assistant helps move the conversation further along in the writing process. Students take ownership over revision, with more informed and thoughtful questions about passages of their text, and motivation to iterate and improve. Teachers get to have deeper conversations with students, point them in the right direction, and watch them go. For the Revision Assistant team that has designed and written the feedback engine powering this interaction, that collaboration is exactly what we call success.

5.0 Comparison with Previous Writing Software

Our approach to supporting student writing breaks the status quo of automated writing assessments by focusing on feedback first.

In the past, other automated assessment platforms had expert statisticians and psychologists define the features of "good" writing. They treated finding the right set of features as a puzzle, with a goal of finding generic metrics that could be applied across writing tasks and genres. Simple features, like word count and Flesch-Kincaid reading level, were paired with regressions to boil any text down to a few dozen universal measurements. The actual content of a student's writing was ignored or simplified to a more computationally efficient representation.

These expert-driven approaches made sense in the 1990s, when those algorithms were invented. They were less complex and were a better fit for the power and speed of contemporary computer processors. However, while they can score summative tasks like standardized tests effectively, such universal evaluation methods offer little value as formative tools for teaching and learning. They lack the flexibility to adapt and align to new standards and curricula. By ignoring content and style, these approaches also miss the chance to highlight specific choices students make in their writing. Thus, the only feedback such legacy systems can offer is overall feedback on entire documents. Our research has consistently shown that this kind of general feedback does not motivate students to revise and improve as effectively as specific, targeted feedback can.

In contrast, Revision Assistant is designed to change students' writing practices. Our bottom-up, data-driven approach to identifying features of strong student writing gives us a window into not just what final score a student receives, but which specific combinations of words, phrases, and passages in their writing led to that score. By generating targeted feedback that uses this information, students receive clear next steps for revision and improvement.

The companion piece to this white paper shows that our approach works. An analysis of nearly 3,500 students' drafts from Revision Assistant classroom pilots shows that students write more, revise more, and score higher when writing with Revision Assistant. It can be found online at go.turnitin.com/ra-pilot-study.

As we continue to study Revision Assistant's implementation in real classrooms, our research team will continue to study best practices on automated writing feedback for writing and share what we learn with the education community.

5.1 Conclusion

Supporting writing with automated feedback is an iterative process—it will never be complete. We will always go between the classroom and the code. As we see students building writing skills with the feedback Revision Assistant provides, we will do more of what's working, and we'll update the approaches that aren't as helpful. As we implement new ideas, we'll go first to the schools using Revision Assistant and watch to see what gets teachers excited. They're the ones who know what works.

We believe that we've made progress on this challenge—but only by using the best algorithms and partnering with dedicated educators that can provide expertise in assessment and content development. We're excited to keep up the work!

Appendix A
Deep Dive: How Revision Assistant Discovers and Analyzes Specific Features of Students' Writing

"Call me Ishmael."

This is three words—not much to go on, right?

Words capture vocabulary and topic.

In our system, each individual word of a text is catalogued separately. In this short sentence, we'd record the use of "call," "Ishmael," and "me" (as a first pass, we won't look at the structure of sentences, so I've listed the observed vocabulary in alphabetical order, rather than the order it appears). These first three features gives us a brief sense of the content we're about to see.

Phrases clue us in to subject matter and discourse.

We then look at paired collocations of adjacent words. "Call me" and "me Ishmael" are tracked, as is "Ishmael." This gives us another three features, but at a slightly more connected level, looking at pairs of words next to each other. This gives us a sense not just for what content the student used but how they used it in the context of phrases and sentences.

Parts of speech show us style and structure.

Next, we can abstract away from the specific words that a student used, and think about the parts of speech that they used to construct a sentence. In most English classes, students are taught a small number of these parts of speech, typically eight (nouns, verbs, and so on). In modern linguistics, there are far more categories than this, over forty in all.

This sentence gets recorded as:

BOL	Present-tense-verb	Personal-pronoun	Proper-noun	Punctuation	EOL
	Call	***me***	***Ishamael***	***.***	

We can now apply the same process of looking at paired adjacent words, but at the more abstract, syntactic level. This gives our software of looking beyond the specific words that a student used, and look instead at the style and phrasing of their sentences. You'll also note that we added tags at the start and end of the sentence. The ends of a sentence have a special power in writing—they're highly important to how a sentence is constructed. When we look at this syntactic level and attempt to capture style, noticing how a writer starts and ends their sentence gives us significant information that we can't ignore. In total, if we look at pairs of adjacent part-of-speech tags, this gives us another five features.

Letters break things down when words aren't enough.

We can also break down a sentence below the level of individual words, and look at just the letters that students are writing. In this case, we can take a microscope to words and look at them as pieces, like "Ishm" at the start of "Ishmael." This gives us several more features (thirteen more if we use blocks of four characters and count spaces).

This might seem trivial, but it isn't. Shared word forms (roots, endings, and etymologies) can tell a story in combination that might be lost when looking at words in isolation. Memorable turns of phrase like "signed, sealed, delivered" or "Veni, vidi, vici" catch readers attention not just for the words they use, but how they share letters and sounds. Words that share common roots—either the content, like "re**ceive**", "con**ceive**", and "de**ceive**", or the inflection, like "comprehen**sion**", "apprehen**sion**", and "indeci**sion**"—can also shed a light on the writer's choices without relying on exactly matching words. Skilled writers will use these patterns to their advantage, without relying on parroting an exact vocabulary, and it is useful to recognize those patterns.

This same approach works at the other end of the spectrum, too. Signs of a struggling writer—common misspellings, switched vowels, or other kinds of malformed words—often have a tendency to show up at the level of letters, not words. Providing good feedback relies not only on seeing exemplars of the best students, but recognizing and identifying the patterns of the students falling behind, as well.

These features are discovered, not dictated

In just that first three word sentence, the approaches described above have given us a lot of evidence: 24 distinct features. Combing through a full document would give us many thousands more. But we can see now that for most of the elements of a text that a skilled reader would catch, there are equivalent ways to discover those elements with combinations of these features. In a workshop, the writer might receive feedback on the directive tone, or the choice to start the document with an introduction. This is not far removed from recognizing the phrases "call me", the sentence-initial present tense verb (rare outside of directives), or identifying that a proper noun was used in the introduction. No one feature tells a story, but a skilled reader would clearly be able to see the choices made by a writer emerge from these features.

Appendix B
Deep Dive: Using Validators to Ensure Meaningful Feedback

Revision Assistant is designed to provide meaningful feedback to students who are working hard on their writing. It can't, and shouldn't, attempt to give feedback on anything typed into a web browser. Revision Assistant's scoring and feedback are accurate when students' writing is original, on-topic, and long enough to express their ideas; the system offers the following guidance and tools to help students meet this bar.

Keep Writing!

Our feedback engine expects essays that are reasonably complete. If an essay is too short, there won't be enough evidence to make an informed assessment. Revision Assistant cannot offer feedback on any essay with only a handful of sentences, and it doesn't try to. Instead, we inform students when their work is too short to receive accurate evaluation and feedback. We also offer prewriting functionality that allows students to work through their ideas without structuring them as a complete essay draft, and to easily transfer those ideas to the writing space at any time.

Say It In Your Own Words!

In source-based writing prompts, struggling students will sometimes overuse excerpts from source documents instead of presenting ideas in their own words. As text from the source document is usually well-written and on-topic, this has the potential to overestimate a student's score. This can be a teachable moment for students learning about original writing. If too much of an essay is detected as being overtly lifted from source documents, Revision Assistant will return the text to the student with an instruction to compose with their own voice.

Stay Focused!

Our scoring models are trained on prompt-specific training sets of on-topic student essays. Writing on other unrelated topics will include many of the same features that serve as evidence for a particular prompt, but when used in a different context, those features will not be meaningful. A pair of validators work to filter out essays that don't fit our training set.

First, we learn to recognize "topicality" of a text—a measure of which vocabulary that occurs more often in a prompt's training set than in general English writing. This works especially well on tightly-focused writing prompts, where prompt-specific words and phrases are likely to be shared between student essays. Second, we turn our attention to the density of features our scoring models expect. Recall that features are discovered anew for each writing prompt and each training set. We rely having some of those features throughout each essay, but some texts are so densely or sparsely packed as to appear suspicious or artificial. This is a warning sign for us, and Revision Assistant learns to avoid giving feedback on such texts.

In either case, Revision Assistant instructs the student to re-consider what the prompt is asking of them before revising so that the system can provide them feedback that will be meaningful.

About Turnitin

Turnitin is revolutionizing the experience of writing to learn. Turnitin's formative feedback and originality checking services promote critical thinking, ensure academic integrity and help students improve their writing. Turnitin provides instructors with the tools to engage students in the writing process, provide personalized feedback, and assess student progress over time. Turnitin is used by more than 26 million students at 15,000 institutions in 140 countries. Backed by Insight Venture Partners, GIC, Norwest Venture Partners, Lead Edge Capital and Georgian Partners, Turnitin is headquartered in Oakland, Calif., with international offices in Newcastle, U.K., Utrecht, Netherlands and Melbourne, Australia.

sales@turnitin.com
turnitin.com