

# **The Impact of Automated Assessment on Student Outcomes: A Hanguk University Case Study**

**Todd Jobbitt, PhD, Hanguk University; Robin Donaldson, PhD, ecree; Jamey Heit, PhD, ecree**

## **Introduction**

With advances in automated assessment technology, it is now possible to provide students with feedback on their writing in a similar manner as English teachers. In addition to the quality of the feedback, this technology provides students with comments on their writing in real-time. Finally, the availability of automated technology at any time of day provides students with a learning aid that can be accessed on the student's schedule.

These characteristics suggest that automated assessment technology can be an invaluable resource for teachers who ask their students to write as part of their courses. This case study utilized ecree's automated assessment software to determine the extent of the benefits students would realize. Our thesis was that we would see a general trend of improvement in student writing.

To determine impact, we focused on three metrics: how engaged students were with the software, the impact the software had on student outcomes, and the total number of *additional* hours of practice that students had as a result of being able to use the software in conjunction with their course.

## **The Student Experience**

Students were given login credentials that allowed them to access ecree's web-based platform. Once they logged in, students would select the assignment they had completed and upload their draft. In less than a minute, the software would then provide students with feedback on their draft.

The feedback students received emphasized the foundational elements of good writing: statement of purpose, organization, use of evidence, and discussion of that evidence as it related to the student's central claim. Students also received guidance on how to correct spelling and grammar mistakes when needed. Importantly, students using the software *do not see scores* for their essays.

The class included fifty students. These students were undergraduates majoring in international studies. For the study, students were assigned four papers over a term lasting 16 weeks. Use of the automated assessment software from ecree was encouraged but not required

## Results

We discuss the results for each of the assignments and then examine what the data for each assignment says about the three metrics we followed in this study. The following table provides an overview of student outcomes.

	n <sub>1</sub>	n <sub>2</sub>	+/- M (SD)	Improvement/Results		
				(+)	(0)	(-)
Assignment 1	29/50	20/29	2.25/3.83	44.4%	22.2%	33.3%
Assignment 2	27/50	15/27	3.73/6.75	60%	13.3%	26.7%
Assignment 3	28/50	17/28	3.8/2.6	58.8%	11.8%	29.4%
Assignment 4	21/50	13/21	2.75/2.71	30.8%	15.4%	53.8%

Several trends emerge in looking at student outcomes. The first one to note is that student engagement was consistently strong. With the exception of the final assignment<sup>1</sup>, more than half of the class opted to use the automated assessment software to receive feedback on an initial draft (n<sub>1</sub>). Further, roughly 60% of students chose to resubmit at least one additional draft after receiving feedback (n<sub>2</sub>). This level of engagement suggests strongly that students found clear value in the automated assessment's feedback.

As a final comment on student engagement, we would like to highlight that students were not shown their scores in the automated assessment software. The only thing students see is written feedback on how to improve their writing. Given that there is only written feedback available, we can conclude that engagement with the software was the result of student interest in getting the consistent feedback on foundational elements of writing that the software offers.

The second trend we would like to highlight is the consistency with which students *did not* improve. We find these instances to validate the impact of automated assessment software for two key reasons. First, learning is not a process where we see only positive progress. If students are trying to understand a new concept, we should expect students not to improve immediately. The instances where student scores dropped from first to last submission suggest strongly that use of the automated assessment software helps students learn as they would from a human teacher. Declines in scores indicate that students are actually working on the elements that the feedback highlights as needing improvement. Mistakes *should* happen when students are engaged fully.

In examining the data regarding decreases from first to last submissions, we found that 18 students had at least once instance where they submitted multiple drafts and saw a decline in their scores. Of those 18 students, four students had two assignments with a decline. There was one student who declined on three assignments between first and last submission. These data indicate that though students will likely experience a decrease in their score on a single assignment, very few students exhibit a pattern of having scores

---

<sup>1</sup> We posit that the drop in submissions for Assignment 4 indicates that students had "checked out" in thinking about their work.

decrease. This supports our claim that declines show students are learning. A single decline is usually accompanied by scores going up between first and last submissions.

A further point to mention in regards to the decline in scores from first to last submission is that this critical feedback *did not reduce student engagement*. Students were willing to continue engaging with the automated assessment software even when they received results that were not positive. Students will often be “turned off” by negative feedback on their papers, so their willingness to continue using the automated assessment software despite negative feedback suggests the software can help offer critical feedback in a way that is “safe” for students to process.

In addition to strong student engagement, then, the results show that students who use automated assessment software do realize better outcomes. To examine this claim further, we looked at average scores across assignments. The class averages for each assignment are as follows:

Assignment 1: 63.04  
Assignment 2: 74.38  
Assignment 3: 64.25  
Assignment 4: 68.14

Assignment 2 is a clear outlier with respect to class wide performance. As such, we set aside this assignment in our analysis of macro trends in the course (though we note again that within Assignment 2 we found similar use and impact dynamics). Setting aside the outlying assignment, we see a clear trend of improvement from one assignment to the next. The difference in class average from the first assignment to the last assignment is more than five numeric points. Moreover, from one assignment to the next there is steady improvement in the scores.

For our final analysis, we looked at the total number of additional papers the automated assessment allowed students to submit. Because students would otherwise not have had access to the in depth feedback on their writing (students were not allowed to submit drafts to the professor), we take the total number of additional papers as a good proxy for “added pedagogical benefit.”

We calculate that each draft of a paper submitted provides students with 30 minutes of additional engagement on their writing. This figure accounts for the time spend writing a draft and processing the feedback received.

Using this approach, the total additional hours of practice for students in this course was 84.5 hours (169 additional drafts submitted). Given the time pressures teachers face when grading student work, we take this additional practice as a clear indication that automated assessment is a viable educational resource. Simply, there is no other way this class’s students would have had access to this much extra practice on their writing.

## Pedagogical Reflections

It is important to keep in mind the impact that implementing automated essay scoring software may have on students in conjunction with oral and written peer review feedback processes (Liu & Hansen, 2002). This can be especially important when the student population is of mixed ability or in mixed groups, or may have participants from high-context cultures, i.e., may be more homogenous (or emic) in nature than heterogeneous (Liu, 2000) – all of these factors were present in this teacher’s classroom context. This AES grading impact can also be an important element to consider when the final grading curve is split - one strict scale for the homogenous, i.e., Korean students, and a second scale for the heterogeneous, i.e., international students, at this university.

Perhaps not unsurprisingly, the idea of implementing grading software was initially met with skepticism from this student population. Reservations included students’ prior experience with other grading algorithms and a strong preference (reliance?) for instructor marking. The overall perspective of the students could be summed up as, “The instructor really understands what I am trying to say, what I mean, in the paper; the computer doesn’t.” This is not unusual in regard to the prevalence in today’s ELT classroom setting of ‘meaning’ over ‘form’, but that does not mean that the argument always holds weight, something that could be further compared or researched. However, this feedback implies certain pedagogical reactions and preparations.

First, as students are becoming more sophisticated writers in some regard related to the use of certain online writing platforms like blogs, wikis, etc., (Warschauer, 2006), when implementing automated essay scoring electronic peer feedback, it is important to thoroughly train students in both the face-to-face process and the automated scoring process. Specifically, a teacher should fully brief students on 1) what the software is and what it does; 2) how it may or may not be different from past systems that they may have familiarity with; 3) how it can generally benefit the students, and 4) how it can support other classroom processes, such as peer review or journaling.

To illustrate, point by point: (1 & 2) The instructor gave a PPT presentation introducing and explaining the ecree website and software, and how the software differed from the more traditional software grading systems which rely on topical or pattern-based analysis; generally, ecree uses a newer assessment technology called “mechanistic assessment” that gathers essay data points and computes them into metrics that can help measure aspects of quality essay writing and then give necessary feedback to the student; (3) Also, the instructor explained that the software would be used to *support* their efforts in the class most specifically in regard to the improvement of the own individual papers, and not take the place of instructor marking.

These points were especially important, for several reasons. The software allowed the instructor to see how many students used the software, and how often; it became an informal assessment tool for the instructor to gauge overall class participation on the ecree system – an assessment point which can be incorporated into a future syllabus. (4) To support the ecree feedback, the instructor also had students perform two oral peer

reviews of peer drafts, plus write peer-review reflection journals, and complete both pre- and post software use surveys. The oral peer-reviewing was fruitful in some aspects, and less fruitful in others (see Rollinson (2005), for some common reasons), yet the software was at least one further tool to help students gain feedback on their own writing if their peer review groups were less than helpful.

This in-class oral peer reviewing was done for several reasons: 1) to build more personal and social support in a large-class setting; 2) to train students to be responsible for their feedback to peers; 3) to raise students' awareness of how to get and give peer feedback; 4) to expose students to their peers' writing style and to learn new ideas, as noted by Mendonca & Johnson (1994); and, to allow students to reflect on the peer-review and electronic feedback processes, in order to help them determine what was of most use to them in their learning development. One highlight for this instructor was receiving feedback within a student journal where one student stated that his peer group had even cross-compared their personal feedback (that they had given to their peers) against the ecree software, all done outside of class!

## **Conclusion**

Based on the outcomes from this study, we conclude that ecree's automated assessment software is a valuable learning resource for our students. At a time when teachers are under increasing time pressure and outcomes are more important than every, the results in the study should provide a measure of confidence that education can be transformed if we are willing to view technology as a partner in our higher education system.

We understand the concerns both students and teachers have about the idea of automated assessment technology. In the end, however, we cannot ignore two clear takeaways from this study: (a) students consistently engaged the software voluntarily and (b) engagement led to demonstrably better outcomes.

The data are clear that ecree's software is not a magic bullet that will make teachers irrelevant. Rather, the software is another resource that helps students gain extra practice. This practice leads to stronger foundational skills. We were glad to see that students' scores would go down on occasion because this is a strong indicator that students are struggling to master important concepts.

Given the strong engagement and impact we see in this case study, we suggest one way to incorporate ecree's software into the classroom is to create low stakes assignments. Students can (and will!) use the software for extra practice. If students were given a bank of additional assignments and asked to complete a certain number as part of their classroom participation grade, additional practice would be generated without demanding more of a teacher's time. The extra practice students had during this study compounded with additional practice should create a virtuous learning cycle. More frequent practice and additional feedback should strengthen students' grasp of fundamental writing skills.

On the teacher side, we suggest that clear benefits will follow the use of ecree's software. When teachers are not spending their limited time focusing on matters that can be addressed through ecree, they will have more time and energy engaging students on conceptual matters.

In closing, we see these results as an indication that ecree offers a scalable virtual partner for universities. The impact of the software on student engagement and outcomes creates greater efficiencies for the teacher and students. Better outcomes and better engagement will provide a better learning experience overall for all stakeholders.

## References

Hyland, f. (2000). ESL writers and feedback: Giving more autonomy to students.

*Language Teaching Research*, 4, 33-54.

Liu, J. & Hansen, J. (2002). *Peer response in second language writing classrooms*. Ann

Arbor, MI: The University of Michigan Press

Mendonca, C. & Johnson, K. (1994). Peer review negotiations: Revision activities in ESL

writing instruction. *TESOL Quarterly*, 28, 745-769

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal*, Vol.

59(1).

Warschauer, M. (2006). Invited commentary: New tools for teaching writing. *Language*

*Learning & Technology*, 14(1), 3-8.