12-2017

Reduce Your G: Evaluating and Optimizing the Use of Quantitative Feedback for Writing Students

Randy Laist
Goodwin College, rlaist@goodwin.edu

Reduce Your G:

Evaluating and Optimizing the Use of Quantitative Feedback for Writing Students

Randy Laist

Goodwin College

Author Note

Correspondence concerning this article should be addressed to Randy Laist, Goodwin College, 1 Riverside Drive, East Hartford, CT 06118. E-mail: rlaist@goodwin.edu

Reduce Your G:

Evaluating and Optimizing the Use of Quantitative Feedback for Writing Students

Abstract

The purpose of this study is to evaluate whether providing writing students with numerical feedback regarding the frequency of technical errors in their writing assignments can be a useful tool for reducing the number of such errors in subsequent assignments.

Student essays from four online sections of a 300-level writing class were scored using a system that resulted in a "g-score," or "glitch" score, a simple average of the amount of errors per 100 words in the essay, as calculated using a codified system. In the control group, the students were not notified of their g-score, and in the first experimental group, they were. In the second experimental group, students were asked to reflect in general terms on strategies for reducing their g-score, and in a fourth experimental group, students were asked to reflect on the same topic in more specific terms.

Statement of the Problem

College-level composition instruction encompasses a wide variety of learning objectives. In addition to cultivating critical thinking skills, learning rhetorical strategies, and pursuing other higher-order objectives, there is also an expectation that writing classes will produce students who are more likely to write in standard, error-free English. For the last several decades, the consensus among educational researchers has been that decontextualized grammar instruction has a negative effect on student's ability to write well (Braddock, Lloyd-Jones, & Schoer, 1963; Micciche 2004). Over the years, educational researchers have proposed a number of strategies for teaching grammar "in-context," or using holistic strategies that offer other alternatives to "traditional" grammar instruction (Lane, 1993; Weaver, 1998; Schuster,

2003). Despite these innovations, however, "traditional" grammar instruction remains widespread (Jean & Simard, 2011), and, as Lunsford and Lunford (2006) discovered, the actual number of errors made by college students has remained remarkably consistent over the last 100 years.

Lunsford and Lunsford (2006) reached this conclusion by using a codified system to rate the number of technical errors that student writers made and comparing their results with past research. According to their research, the average first-year college student commits 2.45 errors every 100 words. In the 10 years since Lunsford and Lunsford's study, self-quantification has been an increasing trend, enabled by new technologies like FitBit and personal analytic services such as Daytum and Rescue Time.

The potential benefits of a "Quantified Self" (QS) approach have been discussed in relation to various fields, including health and fitness (Almalki, Martin-Sanchez, & Gray, 2013), psychological and biological science (Swan, 2013), financial work (Ke & Shi, 2014), and Educational Technology (Lee, 2013).

To the best of my knowledge, however, a QS approach has not been systematically applied to increasing the accuracy of student writing. The nearest examples of a widespread practice that applies quantitative methods to student writing focus on English as a Second Language learners, such as the use of error analysis (Richards, 1984) and Computer-Assisted Language Learning (Lim, 2010). The method used by Lunsford and Lunsford (2006), however, provides a simple, intuitive model for quantifying a mainstream writing students' technical proficiency. Providing students with a simple tally of their average rate of errors, it has been proposed, may "support reflective learning practices" (Rivera-Pelayo, Zacharias, Muller, &

Braun, 2012, p. 114) for students who are working to hone a professional-sounding writing voice.

Method

The study examined essays written by students in four online sections of a 300-level undergraduate online class called "Advanced Writing" at Goodwin College in East Hartford, Connecticut. The online content of the class is standardized, so students in both the control and experimental groups were exposed to exactly the same instructional material. In all four classes, the instructor identified technical errors in student writing through the use of editor's marks written into the digital file of their assignment. The only systematic difference among the four sections was the inclusion in the experimental classes of the students' g-score which was provided alongside their official letter grade for the assignment.

G-score

The error frequency quotient was referred to as a "g-score" to make it sound more fun and less intimidating than a more "scientific" sounding name - for example, EFQ (Error Frequency Quotient) - might have made it appear. The formula for calculating g is as follows:

$$G = \text{error rate} / \text{word count} \times 100$$

The definition of Error Rate was delimited to make it as objective as possible. An error only contributes to the g-score if it is objectively "wrong" by any editorial standard.

- Error rate includes:
    - Misspellings
    - Agreement errors
    - Sentence structure flaws (fragments and run-ons)
    - End-punctuation

- Does not include:
    - Formatting and citation errors
    - Serial commas, commas between independent clauses joined by a conjunction, short introductory sentence elements
    - "Rhetorical" fragments

The Word Count is based on the text of the student essay itself, and does not include front matter or references.

The experiments

The control group, a class of 21 students, represented the typical approach to providing online feedback to students about their writing. Students received their edits on their written assignments, along with a letter grade calculated according to a rubric that included technical considerations as only one of three main categories of assessment (a rubric which also comprised evaluations of content and style). Students were encouraged to review the professor's edits with the following announcement:

- Your Week One assignments have been graded. If you want to do well in this class and improve your writing, your best bet is to pay careful attention to the kinds of corrections I make on your papers, and to try to avoid making the same mistakes in your next writing assignment. Keep at this for seven weeks, and you will certainly see improvements in your writing.

The instructor calculated student g-scores but did not disclose them to the students.

Experiment 1. In addition to their official grade, students also received a g-score, along with an incentive to improve. An announcement informed the students that,

- Students who achieve a dramatic reduction in the g-score on their second assignment (>1 point or >50%) will receive five bonus points on their Rhetorical Analysis essay. The best way to decrease your "G" is to carefully review the errors I marked on your assignment, and then to work to avoid those same errors as you proofread the next assignment.

Experiment 2. Students received a g-score and the identical incentive to improve, and were also required to write a discussion board post articulating a strategy for lowering their g-score.

- After Week 1: Use this link to post a comment about what you will do to try to lower your G-score for the Week 2 assignment.
- After Week 2: Use this discussion board to reflect on your G-score for the Week 2 assignment. Did it go up or down? What kinds of mistakes contributed to your Week 2 G-score, and how might you go about reducing your G-score for the next written assignment?

Experiment 3. Students received a g-score and the incentive to improve, and were also required to identify and discuss specific errors in their writing.

- In this discussion thread, please transcribe three sentences that contain errors that I corrected in your essay. For each of your three sentences, please provide a brief description of the error you made, a note about why you made it, and a plan for how you will try to avoid making similar errors on the next assignments.

Results

The baseline results indicate that, in the typical instructional model, students in the control group (n=21) made more errors on average in both the second and third assignments than they did on the first, with a sharp spike in the second assignments and an average increase in

their g-scores from the first to the third assignment of 4.86% (as calculated by hand) (see Figure 1). In experiment 1 (n=20), we also see the spike in the second assignment, but there is slight indication of minor improvement between the first and third assignments, a decrease of 0.52% (Figure 2). Experiment 2 (n=23) follows the same spiked pattern as the other two charts, but now we see slightly further improvement from the first to the third week, a decrease in average g-score of 2.38% (Figure 3). In Experiment 3 (n=20), however, we see a strikingly different pattern emerge, one showing an undeniably significant drop for each assignment, -16.56% from assignment 1 to assignment 2, and -21.03% from assignment 2 to assignment 3, constituting a total reduction in g-score by 34.11% from the first to the third assignments.

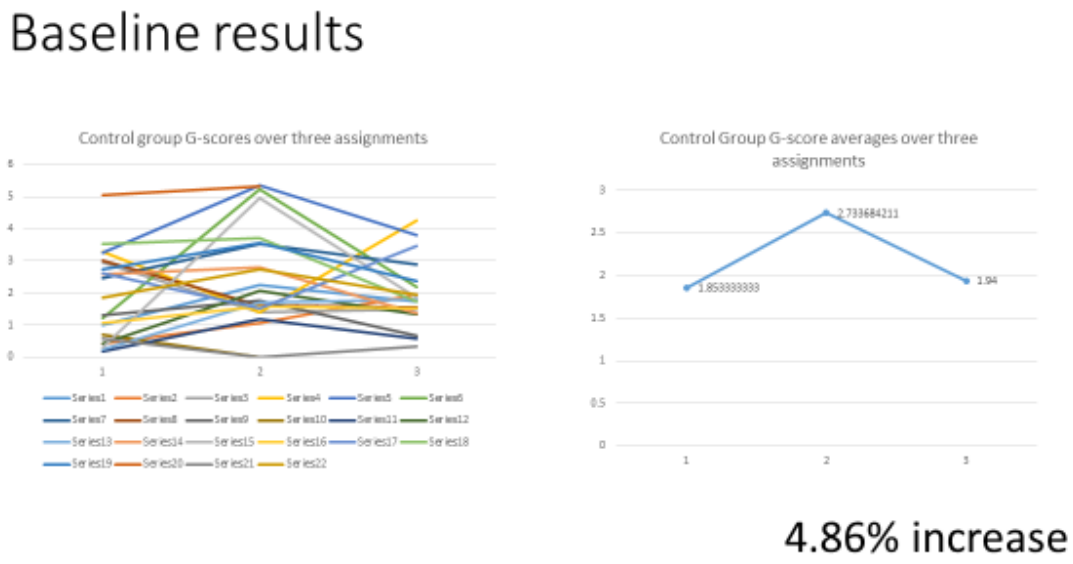

Figure 1: Results from the control group

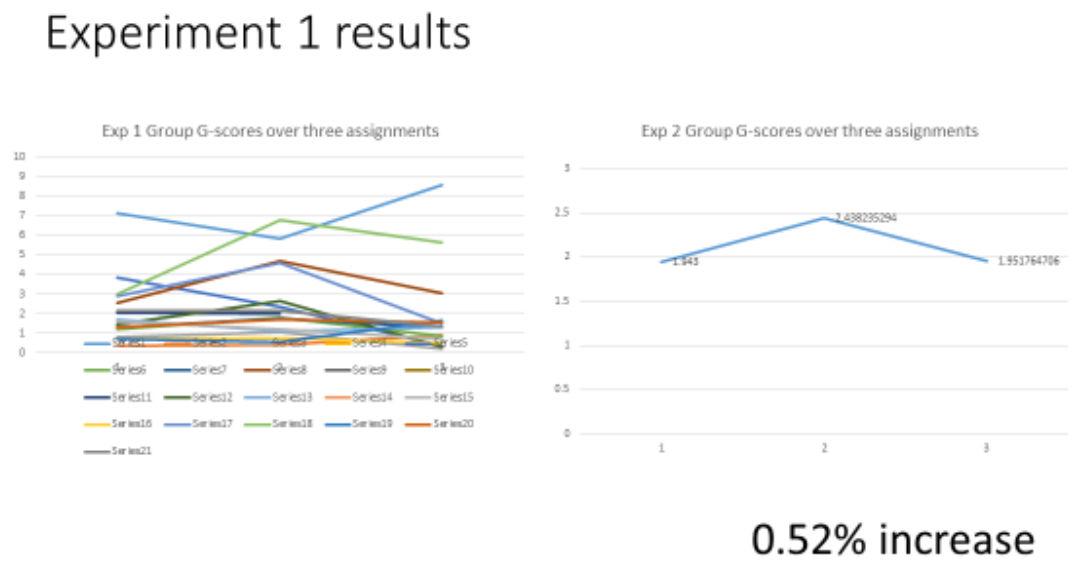

Figure 2: Results from the first experimental group

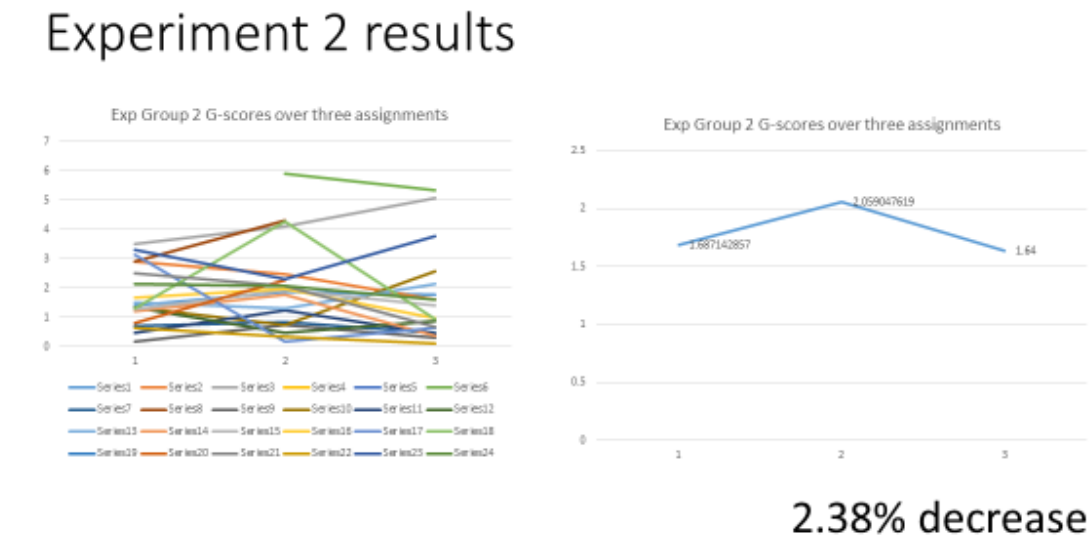

Figure 3: Results from the second experimental group

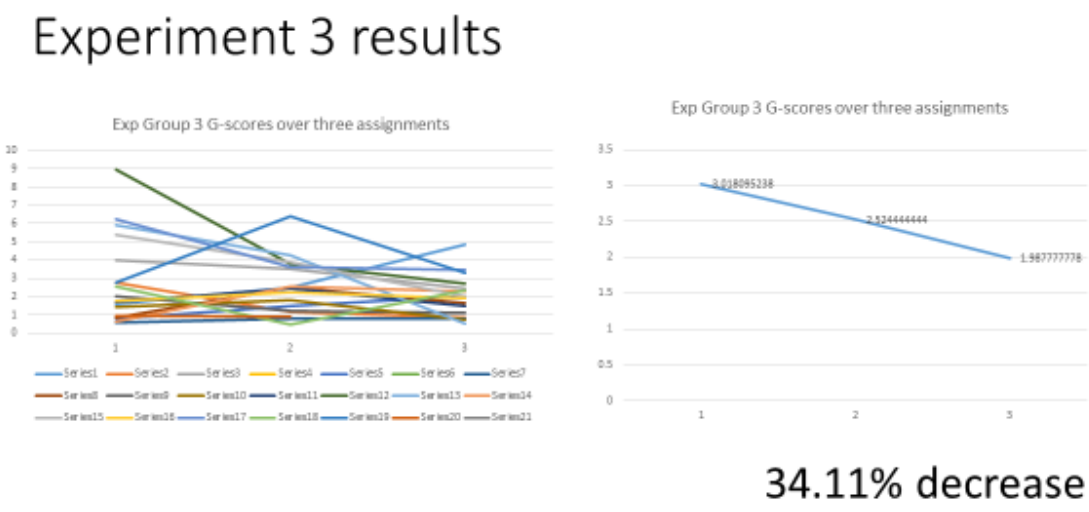

Figure 4: Results from the third experimental group

Discussion

It turns out that the most prominent outcome of calculating students' g-scores has not been the motivational force of the g-score in influencing student error rates, but the information it has provided to the instructor about the efficacy of various instructional strategies.

It should be acknowledged that this study comprises certain design flaws that qualify the spirit in which the results should be considered. Most importantly, the study included no procedure for "blinding" the experiment. Combined with the unavoidable subjectivity involved in calculating the g-score, this makes this experiment extremely vulnerable to the charge that experimenter bias may have influenced the results. My only response to this charge is that, as a researcher, I often felt that I was actually discovering something, rather than simply inventing

evidence for my preconceptions. I was surprised, in the first three groups, to see the G-scores rise in the second week, and I was surprised to find that they didn't go down at all by the third assignment. I was also surprised to see how high the initial score was for Group 4, and I was frankly surprised to see it go down so dramatically. The dramatic difference in the initial scores for Group 4 also present a problem for the validity of the experiment. It is possible that the scores came down in group 4 simply because they were so elevated to begin with. It will be necessary to attempt to replicate the experiment with a more typical group of students. In any case, these findings primarily represent an opportunity to reflect, rather than conclusive evidence of anything.

One peculiar finding was the consistency of the second week "bump" in the control group and first two experimental groups. This phenomenon may be related to the different writing genres assigned each week. The first assignment is a memoir essay, which encourages students to write in their own voice, whereas the next two assignments present students with the challenge of expressing themselves in a "professional" or "academic" idiom that may cause them to err with greater frequency. It also may be the case, however, that students typically "set out to impress," putting their best foot forward in the first week, and then drawing back their efforts in the second week as they settle into the routine of the semester. A future study will have to look at this question more broadly in a range of academic settings.

More alarmingly, the most notable finding in the g-scores from the control group was the observation that g-scores did not go down at all over the course of the three assignments, but actually went up, an astonishing 47.57% in the second assignment, and then a more modest 4.86% by the third. The curriculum of the class covers many other topics related to writing - structure, voice, research, and style, for example – but as far as this comparatively quantifiable

variable of technical accuracy goes, it could be said that their participation in the class was having the effect of degrading students' abilities, rather than improving them. These results are actually consistent with research that has found that teacher feedback on student writing has very little effect on improving student outcomes (Hillocks, 1986; Robb, Ross, & Shortreed, 1986). Similar to the second assignment bump, the comparative error rates may be a result of the different writing genres assigned each week, but the control group data undeniably provoke the question of how to provide editorial feedback that incentivizes students to reduce their error rates.

The typical instructional model of pointing out student errors and instructing them, in effect, to "go and sin no more" is decidedly ineffective, at least when it comes to reducing the rate of grammatical errors over a series of three writing assignments. This experiment began with the hypothesis that simply marking student errors provided students simultaneously with too much feedback – there is not a coherent message that pinpoints the meaning of all these marks – and not enough – it does not tell a student what they need to do to improve, or even whether they need to do anything. The first experiment proposed that providing students with a g-score and an incentive structure would have the same effect on students that wearing a FitBit has on my colleagues: provide hard data that sets a benchmark for improvement. Students would have a clear, numerical indicator of their technical proficiency, and, motivated by the 5-point incentive, they would be able to apply their own strategies for moving that number down. Although g-scores for this group improved slightly from the first to third assignment, the shape of the data for Experiment 1 is essentially identical to that for the control group. These data suggest that providing students with g-score data by itself had little if any effect on student outcomes. On average, student papers in the first group showed a higher average error rate on the third

assignment than on the first, although the rate of decline was slightly less than the corresponding decline in the control group.

The second experiment attempted to encourage students to actively engage in a discussion about their g-score. The discussion board feature in online classes is an ideal forum for reflective self-assessment (Kayla & Weller, 2007) and students used this space to reflect on their g-scores and articulate strategies for reducing them. In their discussion board posts, students described proofreading strategies, such as printing out their rough draft, reading their papers out loud, waiting between the writing stage and the revision phase, and clarifying their comprehension of their subject.

- One way I am going to improve my G-score is actually printing out a rough draft and going over it on paper. It may sound counter-productive but, I believe acquiring a written document to go over is more beneficial then a computer screen at times. Also reading out loud has helped dramatically!
- After re-reading my paper I am going to try and lower my G-score by waiting and proof reading my paper. I know I tend to do this alot and proofread after I write it and to me it sounds ok because I just wrote it. A professor told me once to wait a bit because you will find more errors after you take a break from it and it doesn't sound familiar to you. I also need to remember that I need to make sure I am clear on subjects that I am referring to. I think reading out loud will be also beneficial to me.
- After reviewing my first assignment. I realize there are some steps I could definitely take to lower my G-score. Next time I will be reading over my assignment as much as possible to come across any grammatical errors. I will be utilizing my punctuation marks

where needed. I will also be expressing myself more instead of holding back information that is necessary to make my writing superb.

Some research has suggested that cultivating this kind of active reflection on proofreading strategies results in improvements in students' technical proficiency (Abbott, 2016; Turbill, 2000), but the present study found only a minor effect on student outcomes. It is notable that this is the first group that consistently reduced their average g-score from the first to the third assignment, but the shape of the data remains consistent. In experiments one and two, increasing instructional emphasis is being placed on mechanics as a result of the g-score discussion, and this change in the pedagogical climate can readily account for the modest effect measurable among the g-scores in these groups.

With a minor adaptation of the experiment 2 scenario, however, a startling effect appears in the data. For the first time, we observe a substantial decline in students' error rates over the course of all three assignments. The sole difference between experiments 2 and 3 was that, in the third experiment, students were instructed to write discussion board posts that identified specific errors that had been marked on their drafts, to comment on the specific nature of the error they made, and to suggest a strategy for avoiding these errors in the future.

As a result of this innovation, several developments arose. For one thing, it became clear that many students had difficulty reading their edits in Blackboard. Cohen (1987) observed that students typically do not read teachers' feedback on their written work, and this tendency is arguably amplified in a digital climate that does not even require students to crumple up and dispose of a physical piece of paper; they may never even open the file containing the instructor's editorial suggestions. This problem is further aggravated by the technical limitations in the current iteration of online editing software used in platforms such as Blackboard. I have

been teaching with this platform for several years, but it was not until I required students to demonstrate that they had in fact read my edits that I was compelled to reckon with the problem that many students, by no fault of their own, could literally not read the edits I had marked on their essays. Our g-score discussion thread allowed me to respond to these concerns, as can be seen in the following exchange:

> Student: Thank you for the clarification but just as all of the other students mentioned it is very hard to see your correction that you made on blackboard. I tried making it bigger and even printing it but it was still very hard to see the corrections.
>
> Instructor: I sympathize with you about the difficulty reading the edits in Blackboard. Here are a few suggestions that may help:
>
> - Use Firefox or Chrome, not Internet Explorer, to access Blackboard.
> - Don't enlarge the page, it actually screws up the positioning of the edit marks.
> - If there is a mystery about how my edit mark lines up with your text, try to use critical thinking to solve the mystery of what edits go where. This may require some detective work, but it may also be a good way of teaching yourself some advanced editing skills.
>
> Someday soon, the technology will be much better. For now, we have to muddle through as best we can, and try to turn these challenges into learning opportunities.

When they could read the edits, moreover, students often lacked a clear understanding of why the edit had been suggested. The presence of the g-score thread allowed students the opportunity to reveal their misunderstanding, and for the instructor to respond with clarifying instruction, as in the following exchange.

Student: If it wasn't for him I don't know if I would have had the strength to overcome the days and nights of my life; the nights that i could not sleep because i was in pain. I was to have a lowercase "t" instead of an upper case "t"..... again this was a mistake that was accidently made for not reading it out loud.

Instructor: You are right that these are good examples of cases where reading your essay out loud probably would have helped you to catch these errors. The one clarification I would mention is that, in your second example, the problem is not that you need an upper-case T, the problem is that you have a semicolon where a comma would be more appropriate. Here's the deal with semicolons: only use them when you could also use a period. A semicolon is like a "soft" period; it always separates two independent clauses - two grammatical elements that could stand alone by themselves as complete sentences. If you don't have a complete sentence on both sides, don't use the semicolon. Once you know that, you'll never mis-use a semicolon again! See page 590 of your book for further clarification on this point.

Also, everyone should get into the habit of always capitalizing the personal pronoun "I."

In addition to addressing these problems of illegibility and miscomprehension, the G-score discussion allowed students to reflect on particular errors in their draft, and, whereas the g-score discussion in experiment 2 consisted largely of vague platitudes about proofreading strategies, the g-score discussion in experiment 3 delved into very specific issues of mechanics and grammar. Representative student comments focus on questions of sentence structure and punctuation.

- This day was like any normal day same routine from Monday through Friday. The correction: This day was like any normal day. It was the same routine from Monday

through Friday. I had provided two sentences in one. Short sentences are sometimes better than a long sentence.

- Time was moving quickly going through my daily work assignments and answering internal/external calls. The correction: Time was moving quickly, as I went through my daily work assignments and answering internal/external calls. I needed to take ownership of this sentence. I did not provide a subject.
- "I continue watching and was amazed by the energy and strength she had for such a small girl."
- My original sentence for the above was "I continue watching and her energy and strength for such a small girl I was instantly hooked and she became someone I wanted to be one day." I can see now reading the corrected version of this sentence that mine didn't rmake much sense. The wording was out of order and once again I tried to make a sentence longer than it needed to be.
- We were told it was a five hour procedure, which didn't include the two or more hours we would have to wait for her to wake up from the anesthesia. I added a semi colon instead of a comma. I think I made this mistake because I was worried about the sentence being a little too long. To avoid this in the future, I need to study when and where a semi colon should be used.

These comments allowed the instructor, in turn, to provide responses tailored to the context of individual students' technical questions about their own sentences. The g-score component of the class made it possible to have these discussions about technical aspects of writing in a class that is primarily focused on the *content* of students' writing, but which also strives to develop writers who are technically proficient. As a result of these discussions, the

mechanics of writing became an active part of the curriculum, rather than something silently embedded in the shadow curriculum. In fact, when one observes the depth and specificity with which the students reflect on their mistakes in the g-score discussion thread, it is less surprising to see the chart showing the rapid decline in error rates. While it seems that students' receipt of a g-score did not have much of an impact on reducing student error rates, the attempt to quantify students' technical proficiency did result in meaningful insights about how to make editing feedback meaningful. Particularly, this experiment suggests, it is crucial not merely to provide feedback to students, but to invite them into a discussion about the feedback they receive.

The next stage of this assignment, consequently, will invite students to reflect on the instructor's edits, but will not appraise them of their g-scores. This experiment will allow us to see if the g-score itself has any impact at all on student outcomes. As previously mentioned, a blind version of this experiment will be arranged, should circumstances permit. Future research might also inquire into the characteristic g-score statistics of different genres of writing, different stages in students' academic careers, and different disciplines, as well as in non-academic settings.

It would certainly make efforts to employ g-score statistics more practicable if educational and/or word processing software were capable of calculating g-scores automatically. It is easy to imagine that, in the near future, an automated program will provide detailed statistics, not only about the rates of sentence-level error in particular pieces of student writing, but also about the nature of students' errors. Such feedback would obviously allow writing instructors to target their instruction to students' specific needs, and it would facilitate the ability of writing students to pursue their own attempts to eliminate technical errors from their writing. It is also necessary to continue working with online learning management systems

such as Blackboard to refine their technology to make teacher feedback more legible and useful. The experiment also pushes us in the opposite direction, however, suggesting the limits of quantified data in educational settings, a point that has been made recently by popular books such as O'Neil's *Weapons of Math Destruction* (2016) and Abeles's *Beyond Measure* (2015). As motivational as numbers can be, they may be most efficacious as a starting point for inquiry and discussion.

The research concluded that, while providing the students with a g-score did not seem to have any measurable effect on student outcomes, the g-score did provide the instructor with a tool for assessing and revising methods of instruction. A future stage of research will determine whether specific reflection alone is sufficient to reduce students' error rates, even if the instructor does not provide students with their g-scores.

References

Abbott, M. (2016). Learning from errors: The importance of proofreading. *Journal of Nursing Education, 55*(2):119-120. doi: http://dx.doi.org/10.3928/01484834-20160114-13

Abele, V. (2015). *Beyond measure: Rescuing an overscheduled, overtested, underestimated generation.* New York, NY: Simon & Schuster.

Almalki, M., Martin-Sanchez, F., Gray, K. (2013). *Self-quantification: the informatics of personal data management for health and fitness.* Institute for a Broadband-Enabled Society. Retrieved from http://apo.org.au/node/35485

Cohen, A. (1987). Student processing of feedback on their composition. In Wenden, A. and Rubin, J. (Eds.) (1987). *Learner strategies in language learning.* Prentice-Hall International.

Hillocks, G. (1982). The interaction of instruction, teacher comment, and revision in teaching the composing process. *Research in the Teaching of English, 16*(3), 261278.

Kayler, M., & Weller, K. (2007). Pedagogy, self-assessment, and online discussion groups. *Journal of Educational Technology & Society, 10*(1). Retrieved from http://goodwin.idm.oclc.org/login?url=http://search.proquest.com/docview/1287041987?accountid=41699

Ke, M., & Shi, Y. (2014). Big data, big change: In the financial management. *Open Journal of Accounting, 3,* 77-82. doi:10.4236/ojacct.2014.34009

Lee, V. R. (2013). The quantified self (QS) movement and some emerging opportunities for the educational technology field. *ITLS Faculty Publications.* Paper 480.

Lim, M. (2010). *An investigation of students' experiences with a web-based, data-driven writing assistance environment for improving Korean EFL writers' accuracy with English*

grammar and vocabulary (Order No. 3443576). Available from Education Database. (855814401). Retrieved from http://goodwin.idm.oclc.org/login?url=http://search.proquest.com/docview/855814401?accountid=41699

Lunsford, A. A., & Lunsford, K. J. (2008). "Mistakes are a fact of life": A national comparative study. *College Composition and Communication, 59*(4), 781-806. Retrieved from http://www.csudh.edu/ccauthen/575s12/lunsford.pdf

Micciche, L. R. (2004). Making a case for rhetorical grammar. *College Composition and Communication, 55,* 716-737. Retrieved from http://goodwin.idm.oclc.org/login?url=http://search.proquest.com/docview/220689799?accountid=41699

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy.* New York, NY: Crown.

Richards, J. C. (1984). *Error analysis: Perspectives on second language acquisition.* New York, NY: Routledge.

Rivera-Pelayo, V., Zacharias, V., Muller L., & Braun, S. (2012). Applying quantified self approaches to support reflective learning. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 111-114. Retrieved from http://dl.acm.org/citation.cfm?id=2330631

Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly,* 20, 1, 83-93. doi:10.2307/3586390

Swan, M. (2013). The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data,* 1(2):85-99. doi:10.1089/big.2012.0002

Turbill, J. (2000). Developing a spelling conscience. *Language Arts, 77*(3): 209-217. Retrieved from http://goodwin.idm.oclc.org/login?url=http://search.proquest.com/docview/196850907?accountid=41699