

A Google Sheets-Based Approach to Dynamic Written Corrective Feedback: Learning Outcomes and Teacher Time

Andrew J. Barnes*

Waseda University, Japan

Abstract

Part 1: Overview of the Study

Using custom workbooks created in Google Sheets, dynamic written corrective feedback (WCF) was digitized in this study to streamline the workloads of both teachers and students. While previous attempts used Google Sheets for a one-to-one conversion, this study looked to further leverage web-based affordances. With these web-based tools, teachers could more easily take advantage of online resource sharing and consolidate class data using simple formulas. This allowed students' data to be pulled into a master sheet as well as pushing out changes to workbooks. Moreover, Google Sheets' status as a freely available application on multiple platforms, thus not limiting students' choice, was deemed of benefit to this study. However, while the choice of application was integral, the main purpose of this study was to examine the learning outcomes associated with providing different forms of teacher feedback within an adapted dynamic WCF framework. In this study conducted at a Japanese university in Tokyo, the participants were from three classes of Law majors following the same English for academic purposes syllabus over 15 weeks. Group 1 ($N = 28$), acting as a control, received the teacher's feedback only in the form of comments and marginal notes on weekly paragraph writing tasks, but no grammar correction. Group 2 ($N = 24$), in addition to such teacher comments, had all errors underlined and categorised with an error code by their teacher. Group 3 ($N = 19$) received teacher comments and a more direct form of WCF whereby errors were crossed out and the correct form provided alongside an error code. Following the dynamic WCF approach, students were tasked

*Corresponding Author E-mail: a.barnes@fuji.waseda.jp

with correcting multiple weekly drafts, maintaining a tally of errors, and keeping an error log (a more detailed record of grammatical errors made on the weekly assignments). The teacher's task was to provide said feedback and ad hoc grammar lessons based on the class's needs. This was where digitisation through Google Sheets was a boon, allowing the teacher to pool individual students' error tallies together to create a class overview. This assisted with then choosing the ad hoc grammar topics to be covered during class time. However, while the system was streamlined as much as possible through piloting and users' feedback, it raised some potential issues that must also be considered when providing grammar correction and specifically dynamic WCF: that of the teacher's time on task. As such, descriptive statistics regarding learning outcomes are presented alongside the real-world time taken to provide the feedback across the 15-week semester. Finally, the learners' draft completion data are provided for greater context.

Part 2: Relevance to Theme

One outcome of this study is whether the implementation of technology in the classroom does indeed streamline tasks for educators. As such, it touches on an aspect of teacher education that is often missing in the WCF literature, that of the real-world time on task when providing WCF through a digital platform. Finally, it should be noted that the study data were collected prior to the public launch of ChatGPT.

Keywords: Written corrective feedback, Dynamic WCF, L2 writing development, Teacher time

Background to the Study

This PhD study aimed to adapt dynamic written corrective feedback (DWCF) (Hartshorn et al., 2010) using Google Sheets to streamline both the teacher's and students' workloads. With Google Sheets having previously been used for a one-to-one conversion (Eckstein et al., 2020), this study looked to further incorporate web-based affordances. Being web-based allows for easier sharing of resources, the pooling of learners' data within master sheets, and the pushing out of changes to workbooks. However, in light of concerns over a lack of behavioural engagement with digitized written corrective feedback (WCF) viewed outside the Japanese English as a foreign language (EFL) classroom (Barnes, 2023) and the teacher time associated with DWCF (McQuillan, 2012), this paper examined not only DWCF's effectiveness as a tool of second

language (L2) writing development, but also the necessary time cost placed on the teacher providing it.

Literature Review

Dynamic WCF

DWCF is an alternative approach to those found in a more traditional process writing classroom (Hartshorn et al., 2010). In comparative studies, students in a DWCF group have often outperformed an equivalent process writing group in measures of grammatical accuracy (see Ferris & Kurzer (2019) for an overview). However, there is a lack of DWCF studies that include a control group who receive no grammar correction. As a teaching method, it can be described as a combination of the following processes: the drafting and feedback cycle, the tally sheet and error log (both maintained by the student), and targeted grammar lessons. Finally, in terms of grammar correction, it involves comprehensive indirect WCF (underlining plus a metalinguistic error code).

Manageability

A key tenet of the DWCF approach is that of manageability alongside being meaningful, timely, and constant (Hartshorn et al., 2010). However, many of the studies on DWCF, while acknowledging the time cost of providing feedback, did not report any results relating to this. According to Hartshorn, “For the most part, manageability was maintained simply by limiting the new student writing to ten minutes per day” (2008, p. 147). The premise was that a class’s short paragraphs would be easier to handle than their long essays. However, objections have been raised as to whether reported effect sizes for a DWCF treatment would ever be worth the associated time burden placed on the teachers (McQuillan, 2012). Because of this, investigating the overall time cost for administering feedback is of interest.

Research Questions

- RQ1.** Does DWCF significantly improve grammatical accuracy over that of a control group?
- RQ2.** Does DWCF significantly improve other aspects of L2 writing development over that of a control group?
- RQ3.** What is the associated time cost of providing the different forms of teacher feedback?

Methodology

Procedure

Seventy-one Japanese EFL students at a university in Tokyo were recruited from three intact classrooms which formed the experimental groups: control ($N = 28$), indirect ($N = 24$), and direct ($N = 19$)¹. Timed paragraph writing (10 minutes) was performed based on weekly prompts. The following week, the students received teacher feedback in the form of comments and questions. In addition to this, those in the indirect group received underlining of errors alongside a metalinguistic code, while the direct group were further provided with both the correct forms and codes. A weekly redrafting and follow-up feedback cycle commenced until five drafts had been completed or all corrections had been attended to. There were 13 timed writing exercises and 8 drafting cycles in total across one semester, with the students also maintaining tally sheets and error logs. Finally, the teacher/researcher chose and taught the same grammar lessons for all three classes on the basis of the pooled tally sheet data.

Teacher time was tracked from the commencement of providing a class with feedback until completion of all their drafts, with times recorded at the class level each week. The timer was only paused for interruptions and comfort breaks. This method was followed to gain a clearer understanding of the teacher/researcher's associated workload while avoiding a procedure that would increase it.

Representative Results

At the point of the pre-test (Table 1), one-way ANOVA calculations for independent measures revealed no significant differences across any of the indices of L2 writing development at $p < .05$. However, differences in fluency (word count) were approaching significance ($p = .062$). Paired t -tests for independent means revealed significantly higher fluency scores compared with the control for both the direct ($t(37) = 2.04, p = .049$) and indirect ($t(45) = 2.12, p = .040$) groups.

In regards to the post-test results (Table 1), ANOVA tests revealed no significant differences for any of the measures except for the two indices of lexical diversity: moving-average type-token ratio (MATTR) (Covington & McFall, 2010) ($F(2,57) = 5.669, p = .006$) and measure of textual lexical diversity

¹Following exclusions (missing post-tests, dropouts, etc.), 60 students' data were analysed: control ($n = 26$), indirect ($n = 21$), and direct ($n = 13$).

Table 1. Descriptive statistics for L2 writing development

	Pre-test						Post-test					
	Control (n = 26)			Indirect (n = 21)			Direct (n = 13)			Control (n = 26)		
	Mean	SD		Mean	SD		Mean	SD		Mean	SD	
Complexity												
Clause/T-unit	2.652	0.624		2.585	0.622		2.626	0.940		2.908	0.775	
Accuracy												
Error free clause/ clause	0.716	0.144		0.709	0.148		0.752	0.170		0.677	0.145	
Error free T-unit/T-unit	0.385	0.244		0.443	0.235		0.530	0.210		0.365	0.246	
Fluency												
Word count	73.15	27.69		90.57	28.52		96.62	44.08		87.38	25.53	
Word count/T-unit	12.35	2.51		12.64	2.91		12.63	3.12		13.17	3.42	
Lexical diversity												
MATTR	0.681	0.067		0.712	0.056		0.696	0.065		0.680	0.042	
MTLD	42.46	14.58		45.47	12.14		45.08	12.54		42.67	10.70	
Lexical sophistication												
(Token _{adv} × 100)/token	8.388	3.733		8.055	4.532		8.617	2.555		5.735	2.390	
Rhetorical												
appropriateness												
Rubric 0–7	2.539	0.706		2.857	0.727		3.077	0.954		3.115	0.653	
Grammar knowledge												
Score out of 30	22.69	3.27		23.10	2.79		23.54	3.73		23.08	2.19	

*p < .05. MATTR, moving-average type–token ratio; MTLD, measure of textual lexical diversity.

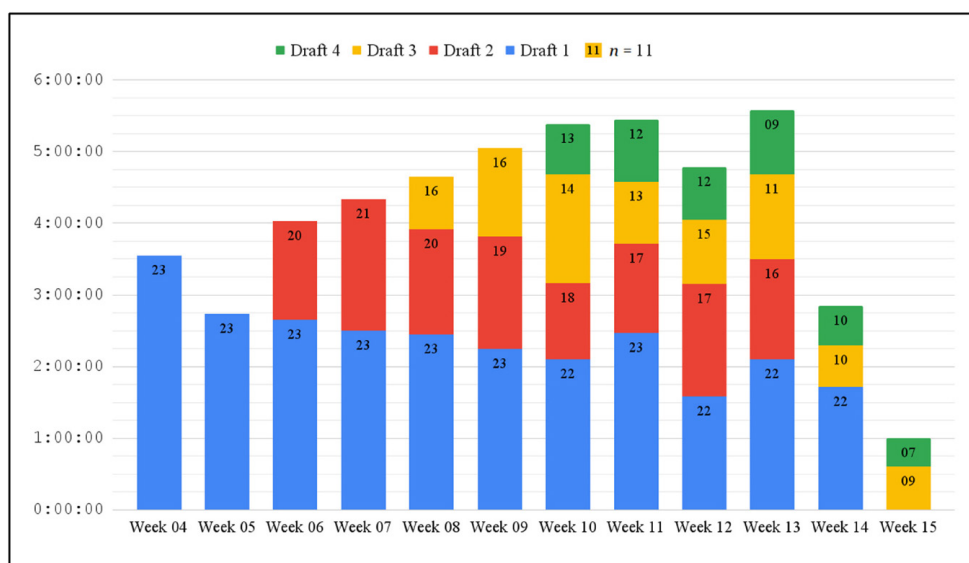


Figure 1. Teacher's time on task by week (Group 2, indirect written corrective feedback, $N = 24$).

(MTLD) (McCarthy & Jarvis, 2010) ($F(2,57) = 3.495$, $p = .037$). Further t -tests revealed that the indirect group had significantly higher scores than the control for both MATTR ($t(45) = 3.305$, $p = .002$) and MTLD ($t(45) = 2.252$, $p = .029$).

Because of space constraints, the full results for the teacher's time on task cannot fully be reported here. However, at the peak of the workload (Week 13), feedback for the control group ($N = 28$) took 3 h 19 m; for the direct group ($N = 19$), it was 2 h 24 m; and for the indirect group ($N = 24$), it was 5 h 35 m, though not all students always submitted drafts. As such, Figure 1 shows a more nuanced view of the associated workload for the indirect class. Finally, the average time spent per student on Draft 1 was estimated as follows: control = 3 m 32 s; direct = 6 m 12 s; indirect = 6 m 11 s. However, in this study, it was the gross time which was of primary interest.

Discussion

RQ1

In response to RQ1, it can be said that comprehensive WCF, be it indirect or direct (with metalinguistic codes), was ineffective. These results run counter to those of previous studies into DWCF (Ferris & Kurzer, 2019). Instead, the findings here were reminiscent of earlier EFL studies in Japan that found comprehensive WCF to be ineffective for improving grammatical accuracy (Robb et al.,

1986). On the other hand, the present study used an adapted form of DWCF, whereas the original methodology may have resulted in positive outcomes.

RQ2

In response to RQ2, the indirect treatment resulted in comparative improvements for lexical diversity when compared with a control group. However, this must be balanced against the comparative improvements in fluency (WC) that the control group made over time. These results are generally in line with previous findings that DWCF did not have a detrimental effect on other aspects of L2 writing development (Ferris & Kurzer, 2019).

RQ3

In response to RQ3, the associated time cost was far beyond the expectations set by a previous study where all first drafts of a class of 15 could be completed within 30 minutes with some practice (Messenger et al., 2020). This may be a result of having a different approach to measuring time, the addition of teacher comments, or the impact of moving from a paper-based-approach to Google Sheets. However, here, the time cost to the teacher was excessive and difficult to justify outside of an experimental setting.

Conclusion

According to the results of this study, only one conclusion can reasonably be drawn within the context of the Japanese EFL classroom, namely that DWCF, as implemented here, was not only ineffective as a trigger for L2 writing development but also placed an extremely large time burden on the teacher providing the feedback. As researchers, we must also consider the reality of teaching when proposing the addition of manual WCF, especially when such a burden seemingly produces comparatively insignificant developmental gains over time.

Acknowledgment

This work was supported by JST SPRING, Grant Number JPMJSP2128.

References

- Barnes, A. J. (2023). Online teacher written corrective feedback outside of class time: (In)effectiveness and task engagement. *Technology in Language Teaching & Learning*, 5(2), 37–58. <https://doi.org/10.29140/tltl.v5n2.911>

- Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type-token ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2), 94–100. <https://doi.org/10.1080/09296171003643098>
- Eckstein, G., Sims, M., & Rohm, L. (2020). Dynamic written corrective feedback among graduate students: The effects of feedback timing. *TESL Canada Journal*, 37(2), 78–102. <https://doi.org/10.18806/tesl.v37i2.1339>
- Ferris, D., & Kurzer, K. (2019). Does error feedback help L2 writers? Latest evidence on the efficacy of written corrective feedback. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: contexts and issues* (2nd ed., pp. 106–124). Cambridge University Press.
- Hartshorn, K. J. (2008). *The effects of manageable corrective feedback on ESL writing accuracy* [Doctoral dissertation, Brigham Young University]. BYU ScholarsArchive. <https://scholarsarchive.byu.edu/etd/1522>
- Hartshorn, K. J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong-Krause, D., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly*, 44(1), 84–109. <https://doi.org/10.5054/tq.2010.213781>
- McCarthy, P. M., & Jarvis, S. (2010). MTL D, voc D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381–392. <https://doi.org/10.3758/brm.42.2.381>
- McQuillan, J. (2012, January 20). All correction, all the time: is written error correction worth the effort? *The Backseat Linguist*. <http://backseatlinguist.com/blog/?p=39>
- Messenger, R. A., Evans, N. W., & Hartshorn, J. K. (2020). Managing dynamic written corrective feedback. *Journal of Response to Writing*, 6(1), 108–138. <https://scholarsarchive.byu.edu/journalrw/vol6/iss1/5>
- Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, 20(1), 83–96. <https://doi.org/10.2307/3586390>