

Technical Report of the Preliminary Results of the Quill Efficacy Trial

Michael Hebert

Pamela Bazis

Julia Roehling

University of Nebraska—Lincoln

Quill Research Study

Quill.org is a non-profit organization that develops online tools to strengthen students' writing skills. This study primarily examined Quill Connect, a tool that employs the evidence-based practice of sentence combining. Quill Connect provides students with opportunities to combine simple kernel sentences into more complex, sophisticated sentences. Students work in Quill Connect independently after completing Quill's diagnostic assessment, which helps identify specific areas of need and recommends appropriate activities in Connect for each student. As students combine sentences in Quill Connect, the program provides immediate feedback and suggestions for improving incorrectly combined sentences (e.g., "try using a joining word, such as and, but, or so"). Students continue revising their combined sentences until they have written a grammatically and logically accurate sentence.

In addition to the Quill Connect tool, which is comprised of practice activities for students to complete at their own pace, Quill has also developed a whole-class learning tool, Quill Lessons. Quill Lessons are designed to be implemented by teachers in an in-person classroom setting prior to assigning specific independent practice activities in Quill Connect. This framework gives teachers the opportunity to provide modeling and guided practice prior to having students practice the skills independently.

This study examines the effectiveness of Quill Connect as a tool for improving the sentence combining skills of students in upper elementary and middle school classrooms. To assure impartiality, researchers at the University of Nebraska—Lincoln (UNL) structured the study as an independent evaluation of Quill Connect's efficacy. The UNL researchers learned

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

about Quill from Quill’s executive director (the “ED”), who contacted the first author of this manuscript requesting a research partnership. To help ensure an unbiased evaluation of Quill Connect, the first author and the ED agreed that Quill would not provide funding for personnel working on the project. However, Quill provided funding for travel expenses related to data collection, which occurred in ten schools in four states. The methods section describes additional safeguards designed to ensure an independent evaluation.

The following research questions drove the design and implementation of this study:

RQ 1: Does instruction in Quill Connect significantly raise student scores on proximal and distal measures of sentence combining?

RQ 2: Do the potential effects of treatment differ based on student level characteristics (e.g., grade)?

RQ 3: Do the potential effects of treatment vary based on dosage (e.g., number of activities completed in the treatment)?

Method

Quill launched in 2014, and many teachers and schools across the United States used the platform prior to this study. Therefore, the first author and the ED agreed to invite schools to participate in the study if (a) the schools planned to use Quill during the 2017-2018 school year, and (b) participating teachers in the school had not previously used the Quill platform. Quill invited ten schools in four states (Massachusetts, Illinois, New York, and Colorado) to participate in the study. Due to practical differences in scheduling and individual school preferences for using Quill Connect alone or in conjunction with Quill Lessons, the design and

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

implementation of the study varied in different schools. The researchers used three different experimental designs within the study with two different implementations of Quill.

Of the ten schools, six (four in New York, one in Illinois, and one in Colorado) elected to use Quill Connect alone as independent practice for students. For the four schools in New York and one school in Colorado, the researchers employed a true experimental design and randomly assigned students to a condition within the classroom (students nested within classrooms). One school (in Illinois) agreed to participate in the study but only had one classroom at each grade level, making classroom-level random assignment within grade levels impossible. In discussions with the school, the researchers also determined it was not feasible to implement student-level random assignment due to practical considerations for teachers. Therefore, no random assignment was conducted, and researchers implemented a simple pre-post design in which all students received treatment.

Alternately, four schools (three in New York and one in Massachusetts) elected to pair Quill Connect practice activities with Quill Lessons. In these schools, random assignment occurred at the classroom level using a quasi-experimental design, meaning classrooms were randomly assigned to treatment or control conditions to facilitate the implementation of the lessons to entire classrooms (classrooms nested within schools).

Due to differences in the experimental design across sites, researchers considered treating these implementations as three individual studies. However, after consulting a statistician on the design and data structure, the researchers determined that a single-level model could be used for the initial analysis.

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

Randomization of Participants and Classrooms

To ensure rigorous implementation of the study design, the first author randomly assigned participants and classrooms. To do this, the first author developed unique ID numbers for all students, classrooms, teachers, and schools, and used a random number table to assign students within classrooms, or assign classrooms within schools, depending on the design for each school. Random assignment was always conducted within classrooms and schools rather than across schools so that a comparable number of participants were included in the treatment and control groups across units of study. The schools and Quill were informed of the results of randomization so that they could implement the study and monitor use of the platform. However, study ID numbers for participants, classrooms, teachers, and schools were never shared with Quill or the schools to ensure independent analysis of the data and maintain confidentiality of the participants following de-identification of the data.

Procedures for Ensuring Independent Evaluation of Quill

Due to the nature of the online log-in features and student/teacher usage of Quill, it was not possible to completely separate Quill from the study. Specific safeguards were put in place to minimize the potential influence Quill might have on the evaluation of their program and to maintain evaluator independence. For transparency, the researchers describe the role of Quill in the research project and safeguards used to reduce the potential for bias in the study from the program developers.

Quill Involvement in the Study. Quill had prior relationships with the schools invited to the study, as those schools had indicated they were interested in using the Quill platform during

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

the upcoming school year. Quill also provided incentives for participation in the form of free access to premium progress monitoring tools. Because Quill requires unique login information for each participant, teacher, and school, Quill necessarily had access to student names, Quill usage data (e.g., number and type of activities completed, minutes spent), Quill's diagnostic assessment information (see Measures section), teacher information, and school information. Teachers also had the ability to log into the Quill system to access reports for individual students in their classrooms.

Quill provided the UNL research team with administrator access to the website to obtain student information. Quill also compiled and shared reports of student usage and diagnostic results with the UNL researchers. To share this data, the researchers used a secure, restricted access, password-protected UNL Box folder. Files including student ID numbers were never stored in the Box folder, so Quill representatives could never link de-identified data back to individual participants.

Quill also provided training to teachers and schools to ensure that they would be able to use the Quill platform. This training is available to schools who adopt the program for use in their classrooms and is therefore consistent with typical implementation. The UNL researchers were not present for the teacher trainings, as this would not be consistent with typical use of the platform. As such, Quill was also responsible for monitoring and prompting use of the Quill program across schools, providing implementation and technical support.

As previously mentioned, Quill also provided travel funding for the researchers to travel to schools for data collection. A contract was established between Quill and the Office of Sponsored Programs at UNL to ensure all funds were used for travel related to data collection

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

purposes only. UNL structured the agreement this way to minimize the potential influence or bias that could occur due to Quill funding personnel.

Research Activities Independent of Quill. The first author obtained Institutional Review Board approval for this study prior to conducting the study. Parents received notification forms informing them of the study and providing them the opportunity to opt out. The researchers also visited each school prior to implementing the study to inform children of the study procedures and obtain verbal assent from the students. Because all schools were already planning to use the Quill program as part of their typical practices, parents and students were notified that students who opted out of the study would still complete the Quill activities as part of typical school activities, but the UNL researchers would not collect or use their information for research purposes. To further promote independence of the data analysis, the research team did not inform Quill which participants opted out of the study.

Aside from usage and evaluation data pulled from the Quill online platform, the researchers directly collected all other participant and school data, including standardized pre-tests and post-tests, demographic information of participants (e.g., gender, ethnicity, home languages), and school-level demographic information (i.e., number of students on free or reduced lunch, number of students with IEPs). The researchers stored this data, and all de-identified data, on a secure UNL server that was not accessible to Quill or the teachers involved in the study. The first author generated a unique ID number for each participant in the study. Only the first author and key study personnel had access to a key linking the names and student ID numbers, as this key was located on a secure server separate from all other study data.

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

The UNL research team, under the direction of the first author, conducted randomization, administration of the standardized outcome measures, scoring of the outcome measures, reliability procedures, data management, and analysis of the results.

Participants

As noted, the researchers drew participants from ten schools in four Midwestern and Eastern states. Seven were urban charter schools, two were urban private schools, and one was a regional public school serving a mix of urban, suburban, and rural populations. After removing students who opted out of the research, the initial sample included 1,141 students enrolled in fourth through eighth grade within 71 classrooms. Each classroom contained a full range of student abilities, and, other than those who opted out, no students were excluded from this study.

Measures

The research team utilized two outcome measures administered as a pre-test before participants received the intervention and then as a post-test immediately following the intervention. Classroom teachers supervised one of the measures, Quill's diagnostic assessment, and administered it through the online Quill platform. Research assistants from the UNL team administered and scored the other measure, the Sentence Combining subtest of the Test of Written Language-4.

Quill Diagnostic Assessment. Quill, the developers of the intervention (Quill Connect), also developed the diagnostic assessment. This tool assesses both basic and advanced sentence structures using 22 questions that cover eight fundamental concepts: fragments, compound

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

objects, adjectives, adverbs, compound sentences, complex sentences, appositives, and parallel structure. Each item on the assessment is scored (1) correct or (0) incorrect depending on whether the student demonstrates mastery of the item's target skill. If a student does not demonstrate mastery of a skill, the student is recommended practice activities in that area. These recommendations are presented to the teacher for review and assignment.

The diagnostic assessment provides a proximal measure of the intervention effects, as the pre-test determined which practice activities were assigned during the intervention and the post-test provided a direct test of the skills taught. The researchers and Quill informed teachers of when their students should complete the diagnostic assessment, ensuring that the pre-test occurred prior to the first instructional session and the post-test occurred approximately 6 weeks later.

TOWL-4 Sentence Combining Subtest. The UNL research team administered the sentence combining subtest of the Test of Written Language-4 (TOWL-4) for students in all schools. This subtest requires participants to read two or more sentences and then combine them into one coherent and grammatically acceptable sentence. This assessment is distal, standardized, and norm-referenced. It was given at pre-test as a measure of student writing ability and at post-test as a distal measure of the impacts of the intervention. There are two forms of the assessment (Form A & Form B), each with 23 items. The assessment increases in the number of kernel sentences to combine for each item, as well as in the complexity of the kernel sentences. Typically, the assessment is administered individually, but the administration instructions also include group administration procedures. Due to the large number of participants, the assessment was administered to entire classrooms of students. Make-up days were built into the schedule for

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

students who were absent on the day of the test. Make-up tests were administered to students individually or in small groups.

The researchers counterbalanced Form A and Form B of the assessments across treatment groups. While typical individualized administration allows for test administrators to stop the assessment for students after three subsequent incorrect responses, this was not possible for group administration. Instead, researchers provided students 45 minutes to complete the assessment. The researchers read the standardized instructions to students, including an example sentence illustrating how to complete the items. Students had the opportunity to ask clarification questions, but test administrators did not answer questions related to specific items.

Graduate research assistants trained in scoring the measure de-identified and then double-scored the assessments. Each example was scored as (1) correct or (0) incorrect. Capitalization and punctuation were not considered in the scoring, and there were multiple correct answers for each item. Agreements and disagreements were identified, and disagreements were resolved through discussion. When consensus could not be reached, the first author made the final decision. The researchers attained an inter-rater reliability score of 95%. Although standard scores and percentiles can be obtained for the measure, the researchers used raw scores for the analyses to better illustrate differences across grade levels.

Instructional Conditions

The first author randomly assigned participants to either the experimental or control condition using the randomization procedures described earlier in the method section, with the exception of one school that did not have random assignment. Students in the experimental

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

condition received either Quill Connect or Quill Connect + Quill Lessons. Students assigned to the control condition did not complete any Quill activities during the study period and instead participated in business-as-usual activities. Control group students were allowed to use Quill following the post-test of the study.

Quill Connect Treatment. The students assigned to the treatment condition received supplemental instruction through Quill's sentence combining program, Quill Connect. This supplemental instruction occurred two to three times per week for approximately six weeks. Quill estimates activities require an average of approximately fifteen minutes per session. The researchers asked teachers to have their students complete 20 activities during the study period. The Quill Connect program contains activities designed to provide direct instruction, feedback, and repeated practice in areas of writing instruction, including fragments, adjectives, adverbs, compound sentences, and complex sentences. Specific activities were assigned to individual students based on their scores on the Quill Diagnostic measure. Each activity included multiple sentence combining items to provide practice in the writing area topic. For each student response, students received immediate computer-generated feedback encouraging them to revise the sentence or confirming that they had written a strong sentence. If the student combined the sentences in an item correctly, the computer-generated feedback would indicate that the student wrote an acceptable sentence with all of the necessary information and provide two or three alternative sentences that would also have been acceptable. If the student provided an incorrect response, the computer-generated feedback indicated the reason the response was incorrect and provided hints for potential solutions, and the students were given the opportunity to revise their response.

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

Control. The students assigned to the control condition did not receive any supplemental instruction in sentence combining through the Quill Connect program. They received the standard English language arts instruction as determined by their school. There was some variety in the instruction provided depending on the school the student attended.

Procedures

Study design guidelines were shared and agreed upon by the research team, Quill staff, participating schools, teachers, and parents. This was a complicated process considering the size of the study and the diverse geographical locations of all involved. Elements included training, pre-assessments, treatment instruction, post-assessments, scoring, and analysis.

Before the start of the intervention, teachers directed all participants in both the treatment and control groups to complete Quill's diagnostic assessment. Within the same week, the treatment group began supplemental instruction using the Quill Connect program.

Instruction for the treatment group took place within the school day at all locations. Teachers were directed to provide time for students in the treatment group to work on the Quill Connect program two or three times per week for fifteen minutes per session. Teachers were also directed to encourage the students to complete approximately 20 activities within that time frame to ensure students had a sufficient amount of time in the program.

At the end of the six to eight week intervention period, research assistants (for the TOWL-4) and classroom teachers (for the Quill Diagnostic) administered the outcome measures to both the treatment and control groups. The classroom teachers directed the students to complete the Quill Diagnostic within the same week of the TOWL-4 administration.

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

Fidelity of Implementation

It was not necessary to collect fidelity information for Quill Connect activities, as the students completed all practice activities independently. The research team also did not have the funding available to go to the schools to collect fidelity information, given the size and scope of the study. Instead, Quill monitored student usage of the program. If students and classrooms were not accessing Quill regularly during the study period, Quill administrators emailed the teachers to remind them to have their students complete 20 activities.

Quill's software collected data on the back-end of the online platform. This data included the number of activities attempted, number of activities completed, number of minutes spent on each activity, and the average number of minutes spent across activities. In some cases, participants forgot to log out of the platform after completing an activity, and the system continued to count minutes until the next log-in period for the student. In those cases, Quill imputed a cap of 30 minutes for the activity (fifteen minutes longer than the amount of time expected for any single activity). Because of this, the number of minutes per activity may be inflated for some students.

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

Results

[Note: The results presented below are preliminary, and were conducted to provide Quill with a technical report on the progress of the analyses. The results do not include data on the Quill Diagnostic Assessment (the proximal measure). Additionally, not all of the assumptions have been checked for the models. Finally, the model used was a single-level model because the Intraclass Correlation Coefficients (ICCs) for the classroom and school levels were .01 and .02, respectively, indicating there is very little variance at those levels. However, the experimental structure for this study was complex, so a single-level analysis may not be the most appropriate model, despite the small ICCs. Therefore, these results should be interpreted with caution, and should only be used for Quill's internal use and to share privately with their board and other stakeholders until all analyses can be conducted appropriately.]

Of the 630 students who participated in the treatment, 178 students (28.1%) completed the full intervention of 20 or more activities. 215 students (34.1%) completed a moderate intervention by completing 10-19 activities. 76 students (12.6%) completed fewer than 5 activities (not including attrition). This indicates low-fidelity of the intervention across teachers and schools for the treatment group, which may not be surprising given the passive nature of the intervention. Teachers may not have scheduled specific time for students to participate in Quill, instead having students participate more sporadically. This seems likely given the number of reminders Quill needed to send to teachers and schools to participate. It may also be that students were not monitored by teachers when asked to access Quill and/or did not complete the activities when assigned. Whatever the reason, the research team elected to conduct the analyses with all

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

students who remained in the sample after attrition, regardless of the number of activities completed.

Analysis of the data indicated that some students ($n = 20$) in the treatment group never accessed Quill for either the diagnostic or treatment activities. It may be that the teachers for these students prioritized other instruction, the students did not comply, or that there was a miscommunication about whether these students would participate in the intervention. Regardless of the reason, the researchers classified this as attrition and dropped such students from the analysis. Similarly, 20 students in the control group accessed Quill and completed activities during the treatment. It may be that teachers of those students did not understand that the students were supposed to be in the control group, or that teachers allowed students to access Quill in a different class (all students were in middle school and changed classes). The researchers classified this as treatment contamination and also dropped these students from the analysis. The resulting sample included 1,101 students, 630 in the treatment group and 471 in the control group.

RQ1: Does instruction in Quill Connect significantly raise student scores on standardized measures of sentence combining?

Results of the single-level regression model are presented below. The researchers elected to remove the Illinois school from the analysis, as all students in that school received treatment. The resulting model examined the effect of the Quill intervention on the TOWL-4 standard score gains with 1,028 participants across nine schools (see Figure 1).

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

Source	SS	df	MS	Number of obs	=	1,028
Model	982.964349	1	982.964349	F(1, 1026)	=	3.90
Residual	258718.307	1,026	252.162093	Prob > F	=	0.0486
				R-squared	=	0.0038
				Adj R-squared	=	0.0028
Total	259701.271	1,027	252.873682	Root MSE	=	15.88

SSgain	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Treatment	1.96258	.9940275	1.97	0.049	.012021	3.913139
_cons	4.067941	.7316939	5.56	0.000	2.632153	5.503728

Figure 1. Single-level regression model examining the effect of treatment with no covariates.

The results indicate that there was a statistically significant effect of treatment. Students in the control group gained 4.07 standard score points on the TOWL-4 sentence combining post-test, and that students in the treatment group gained 6.03 standard score points, an additional 1.96 points over the control group. The effect size for the treatment outcome was 0.16 ($CI = 0.04, 0.17$).

Students	Pre-Test	Post-Test	Change
Control (n=471)	8.36	9.54	1.18
Treatment (n=557)	8.27	9.93	1.66

Figure 2. Raw median scores for students on the TOWL assessment during pre-test and post-test

Students	Pre-Test	Post-Test	Change
Control (n=471)	46.39%	53.99%	7.60
Treatment (n=557)	45.15%	56.79%	11.64

Figure 3. Percentile change on TOWL assessment between pre-test and post-test

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

RQ 2: Do the potential effects of treatment differ based on grade level and Quill Implementation Type (i.e., Lessons or Practice Only) while controlling for pre-test?

Analyses for this research question are on-going. We are still compiling the data for Quill Lessons. The analysis for the complete report is expected in May 2019.

RQ 3: Do the potential effects of treatment vary based on dosage (e.g., number of activities completed)?

Analyses for this research question are on-going. Dosage is being examined in terms of the number of activities completed by students in each condition and number of total minutes using Quill. The analysis for the complete report is expected in May 2019.

Preliminary Conclusions

The preliminary analyses indicated Quill Connect was effective for improving sentence combining outcomes for students in grades 4-8. The resulting effect size for treatment ($ES = 0.16$) is comparable to effect sizes found for other interventions on standardized outcomes of a narrow score (also $ES = 0.16$) as reported by Lipsey et al. (2012) in their review of effect sizes on educational outcomes. This is important, considering that the ESs found by Lipsey et al. (2012) primarily involved effects of instructional interventions, while many of the students in this study completed a practice-only version of Quill with no instruction. Indeed, the additional impact of Quill Lessons indicates that instruction enhanced the effects for students.

One limitation of this study is that many students in the treatment group did not complete the intervention as intended. It is likely that the effects of the study were depressed due to lack of

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

dosage for all the participants in the treatment group. One recommendation is to find ways to motivate teachers and students to use Quill more consistently. The number of students who completed activities may have been even lower if Quill did not remind teachers regularly during the intervention period. It may not be feasible for Quill to provide continuous reminders to all of their users in personalized emails or contacts, so finding other ways to increase usage is necessary. The intervention can only be effective if it is used consistently.

Quill Lessons may be a promising solution. Quill Lessons require more teacher engagement as they are taught by the teacher prior to having students complete the independent practice activities in Quill Connect. Therefore, teachers using Quill Lessons may be more likely to assign activities to students and ensure they complete the practice. It may also have ancillary effects, as teachers who use Quill Lessons more often may strengthen their understanding of sentence combining and help students make connections to it more frequently during other writing tasks. There is no data to support this in the current study. Therefore, the research team recommends that Quill examine this in future research and development of their Quill Lessons product.

Future research examining Quill Connect and Quill Lessons is recommended. While additional analyses of this study are forthcoming, some of these preliminary findings suggest there are specific research questions still of interest. First, a smaller “effectiveness trial” should be conducted with more experimental control to examine the potential of Quill under optimal conditions, including high-quality implementation of the instruction, regular and consistent use of the Quill activities, reteaching of skills when necessary, and regular teacher monitoring of the results with feedback for students. While the current “efficacy study” showed effectiveness, this

PRELIMINARY TECHNICAL REPORT OF THE QUILL EFFICACY TRIAL

was really more of a scale-up study examining the effects of Quill with minimal training for teachers. A study under optimal conditions may provide more information for how lessons could be improved, potential areas for confusion for the students, better ways to help teachers use and interpret the results provided in the platform, etc.

In conclusion, the results of this study are promising. Quill appears to be effective for students in upper elementary and middle grades. More work can be done to refine the instruction and platform and to increase usage. These iterative improvements would be likely to increase the impacts of Quill.