

## Phrasing Feedback to Improve Students' Writing in a Large First-Year Humanities Course

Mairi Cowan, Tyler Evans-Tokaryk, Abdullah Farooqi, Allison Graham, and Michael Kaler

*University of Toronto Mississauga*

Received: 20 April 2021; Accepted: 7 July 2021

On a revise-and-resubmit assignment in a large introductory History course, students were provided with feedback that was phrased either as questions, statements, or imperatives. This study examines which form was most likely to lead to improvement in the students' writing. Generalized estimating equations (GEE) were used to analyze a data set comprising 669 individual pieces of feedback on 67 sets of papers. Researchers found that, overall, students were most likely to implement feedback phrased as imperatives and least likely to implement feedback phrased as questions, and that the likelihood shifted somewhat depending on which aspect of writing was being commented upon; the extent of change required; the students' past performance in the course; and the person providing the feedback.

### INTRODUCTION

Instructors put a lot of time and effort into providing feedback for students. They do this because they recognize that good feedback has the potential not only to justify and explain a grade, but also to help students improve in their future assignments. But that potential cannot be realized if the feedback is not understood, used, or even read by the student. If the feedback is ignored, it becomes merely "dangling data," as Robert Sadler put it (1989). It is worthwhile learning, therefore, what characteristics of feedback are likely to lead to its implementation so that instructors can craft responses to student work that will help to improve the students' future efforts in meaningful ways.

A multitude of factors influence whether or not students make use of the feedback that they receive. Many of these factors, of course, are out of instructors' control. Among those that we as instructors can shape are the form and tone of feedback. Within those parameters, our investigation considers the effectiveness of different kinds of feedback on written assignments. Our most basic research question was whether a student was more likely to implement feedback when that feedback was phrased in the form of a question, a statement, or an imperative. Recognizing that the answer to that question could well differ according to which aspect of writing was being commented upon, the extent of change required, the student's past performance in the course, and the person providing the feedback, we also asked which kind of phrasing was most effective for commenting on a thesis, evidence, organization, and language; if a particular form of feedback was especially effective for a high, medium, and low extent of change; whether a student's grade on an earlier assignment was related to the type of feedback they were most likely to implement; and if the identity of TA made any difference to the implementation of feedback.

In our study, students in a large first-year History course at the University of Toronto Mississauga submitted an essay, received feedback, were instructed to implement this feedback to improve their work, and then resubmitted the revised essay. Using a scoring system in four parts (the format of the feedback, how

well the students implemented the feedback, the extent of the changes requested, and the types of writing problems addressed in the feedback), we considered 1,340 individual instances of feedback (67 students participating in the study; 2 assignments for each student; 10 comments on each assignment). We examined how students' assignments changed between the first and revised submissions, and sought to determine whether feedback constructed as questions, statements, or commands was most likely to prompt improvement.

### BACKGROUND

Many factors shaping students' "readiness-to-engage" with feedback (Handley, Price, & Millar, 2011) are beyond the control, and sometimes even the awareness, of instructors. These factors include students' preconceived ideas about their own strengths and identities (Torres & Anguiano, 2016); their opinions about whether teachers should take effort into account when assigning a grade (Tippin, Lafreniere, & Page, 2012); their perception of how well the grade on their assignment matches their pre-existing standard of what constitutes a good grade (Winstone, Nash, Rowntree, & Parker, 2017) or their pre-determined idea of what grade their assignment should receive (Ryan & Henderson, 2017); their predisposition to either a growth or a fixed mindset when reading feedback (Forsythe & Johnson, 2017); and, if teaching assistants are providing the feedback, the experience of those teaching assistants (Tang & Harrison, 2011; Vardi, 2013).

Overall, researchers have "mixed feelings on the power of feedback to influence learning" (Evans, 2013, p. 73), and faculty do not always follow the advice presented by those who research the principles of effective feedback (Stern & Solomon, 2006). It is clear that the mere existence of feedback does not necessarily mean that students will actually use it (Brown & Glover, 2006). On more specific questions of what makes feedback effective, studies have offered some helpful directions and guidelines for instructors. Clarity in the communication of course objectives and of how assessments fit into these objectives has been shown to improve student learning outcomes. Students improve most when instructors provide clear assessments and frequent feedback (Riddell,

2015), and they appreciate it when feedback is consistent with assignment guidelines (Torres & Anguiano, 2016). In addition to clear instruction, access to exemplars can also help students make use of formative feedback (Sadler, 1989). Students sometimes struggle with reading instructors' handwriting (Hyland, 2000), and may find some colors of ink more likely to provoke anxiety (Davis & Coleman, 2007) although these concerns recede as handwritten feedback decreases, and students often prefer comments that are typed or drawn from a comment bank because they perceive such comments as clearer, easier to read, and fairer (Denton et al., 2008).

Precision in feedback is another factor that can influence implementation, and this is all the more true when set within the context of larger goals for an assignment or a course. Students can feel overwhelmed and intimidated by the amount of feedback on their assignments if there is too much of it (Glover & Brown, 2006; Torres & Anguiano, 2016), and they often prefer precise advice about exactly what to change and how to make improvements (Cho & MacArthur, 2010; Doan 2013; Huxham, 2007; Hyland, 2000; Walker, 2009)—this is particularly the case when the feedback is part of a scaffolded structure or is intended to be used in revision (Walker, 2009). Descriptive feedback demonstrating how to reach correct solutions has been shown to be more effective than simple evaluative feedback (in other words, formative “feedforward” seems more effective than summative feedback); it should address questions of what the assessment goals are, where the student currently is, and the next steps for bridging that gap. In short, feedback needs to be holistically integrated into the course, and the nature of that integration needs to be clear (Bailey & Garner, 2010; Hepplestone & Chikwa, 2014; Lipnevich & Smith, 2009; Mulliner & Tucker, 2017; Robinson, Pope, & Holyoak, 2013; Skinner, 2014; Vardi, 2013). Furthermore, feedback should focus on the task itself rather than on the student who did the task (Tang & Harrison, 2011), and, ideally—from the student's perspective—there should be a balance between constructive feedback (on how to improve their work) and positive feedback (what they did correctly) (Hattie & Timperly, 2007).

Results have been mixed in studies focusing on the relationship between feedback and grades. Some studies indicate that feedback in the form of comments alone results in greater improvement than feedback in the form of comments followed by a mark or in the form of a mark alone (Hattie & Timperly, 2007; Lipnevich & Smith, 2009; Tang & Harrison, 2011). Other research indicates no significant effects of grading on performance and effectiveness of feedback (Dlaska & Krekeler, 2017).

A general theme frequently adopted by instructors is one proposed by Nicol and Macfarlane-Dick (2006), namely, that feedback should be viewed from a student-centred perspective. The kind of feedback that students are most likely to implement, however, does not always overlap with the kind of feedback that instructors provide. Sometimes, this seeming disconnect is for a good pedagogical reason. Although students are more likely to implement feedback that presents easily-achievable objectives, consolidates previous learning, and enhances confidence, instructors often prioritize feedback that aims to develop skills and understanding at a higher level, since this is so important for long-term learning (Donovan, 2014). As suggested above, the appropriateness of different forms of feedback is related to the envisioned goal for that feedback: when assignments are to be

revised and resubmitted, it is all the more important that feedback be targeted and precise (Jonsson, 2012).

## PURPOSE

Researchers have been making increased efforts to assess the impact of academic writing instruction embedded in disciplinary courses, especially in contexts (such as Canada) where separate composition courses are usually not part of higher education curricula (Elliot & Perelman, 2012; Fulwiler, 1988; Kaler & Evans-Tokaryk, 2019; O'Neill, Moore, & Huot, 2009; Pruchnik et al., 2018; Yancey & Huot, 1997). It is already known that students are more likely to use feedback when it is provided on drafts rather than on final assignments (Tang & Harrison, 2011) or on chains of assignments where students take feedback from one stage and apply it to the next (Hounsell, 2007). It is also known that students benefit from repeated assignments of a revise-and-resubmit type (Fisher, Cavanagh, & Bowles, 2011; Freestone, 2009). Our study builds upon this understanding of how students learn from resubmission of assignments, and examines a specific question that we think will be helpful to course instructors, writing specialists, and learning centers in a wide variety of disciplines and institutions: Are students more likely to implement feedback when comments are phrased as questions, as statements, or as imperatives?

Some research has been conducted on the effectiveness of different ways to phrase feedback, but the research has not been done in the context of a large, first-year course. One study that looked at written feedback on reflective essays by pre-clinical students at the University of Groningen's medical school found that phrasing comments as questions was somewhat more effective at stimulating reflection than phrasing comments as statements (Dekker et al., 2013). This finding would seem to link with results from several studies arguing that effectiveness is enhanced by working with the student as a dialogue partner (Dysthe, Lillejord, Wasson et al., 2011; Hyatt, 2005; Mutch, 2003). Two studies examining comments on the writing of ELL students, by contrast, found that imperative comments were particularly effective (Nurmukhamedov & Kim, 2010; Sugita, 2006). Questions, precisely because they are dialogic, require more interpretation and more interaction on the part of the recipient, and may therefore be especially difficult for students with an imperfect grasp of idiomatic English to interpret as the instructor intended. While we do not have demographic information about the specific student population of the course in which this research was conducted, roughly 25% of the students at our campus are international students (Neebar, 2018). Even those students with a level of native proficiency sometimes find it frustrating to be provided with “questions” rather than “answers” (Hewett, 2010).

In this research, our goals were both to bring data produced in a new educational context (a large, first-year History class) into the conversation, and to use these data to explore the effect of the phrasing of feedback. More specifically, we wanted to find out whether phrasing feedback as questions, statements, or imperatives had an effect on the likelihood of students making use of this feedback in a subsequent piece of writing.

## METHODS

All data for this project were collected in the Fall 2018 offering of HIS 101, a first-year History course at the University of Toronto Mississauga that introduces students to world history and discipline-specific research and writing skills. Lectures and tutorials

Paper No.	Q/S/I	A-J	Criterion (thesis, select + present evidence, organization, language)	Extent of change required (low, medium, high)	Score (-1,0,1,2)	Comments
	Q	A	thesis	med		
		B	thesis	high		
		C	thesis	high		
		D	evidence	med		
		E	organization	high		
		F	language	low		
		G	evidence	med		
		H	language	low		

**Figure 1. Rubric for assessing students' implementation of feedback**

were conducted in person, and grading was done online. The major assignment in the course is a set of connected, scaffolded writing exercises worth a total 45% of the final grade. Students submit a tentative Research Question and Thesis Statement (worth 5%) in Week 4 of the course and an Annotated Bibliography (worth 10%) in Week 6. The feedback from these two short assignments prepares them for the Short Research Essay (SRE) due in Week 9, and the Revised Short Research Essay (RSRE) due in Week 12 (the last week of classes). The SRE and RSRE are each worth 15% of the final grade and are the only pieces of writing included in the corpus for this research.

All data collection and analysis methods were approved by the University of Toronto's Research Ethics Board. The participants were students enrolled in seven of thirteen tutorial sections of the course who submitted both an SRE and an RSRE and signed an informed consent form. A total of 67 students met these criteria and were included in the study.

All students in HIS101 were required to submit the SRE in Week 9. After receiving feedback on the SRE in Week 10, students then submitted the RSRE in Week 12 (the last week of class). The SRE was a 2,000-word research essay that made an argument about a primary source selected from a list provided by the course instructor; students were asked to use three secondary sources (approved in their earlier Annotated Bibliography) to support the argument in their SRE. For the RSRE, students were expected to revise the SRE by incorporating both their TAs' feedback on the SRE and writing advice from one of the course texts, *Writing History*. Students were told to identify all changes from the original SRE in bold type.

Students in the participating sections were assigned randomly into three groups. Feedback was provided in question form for Group 1 (e.g., "Where is the thesis statement?"), in statement form for Group 2 (e.g., "The thesis statement needs to be more clear"), and in imperative form for Group 3 (e.g., "Write a clearer thesis statement"). TAs were trained in how to provide feedback in these different ways. They could draw many of their comments from a comment bank created by the instructor in advance and improvised when necessary.

Before the course began, the instructor developed a set of four assessment criteria for the assignment:

1. clarity of the thesis
2. selection and presentation of evidence, including correct citation
3. organization
4. effective use of scholarly language

The TAs put 10 marginal comments on the SRE, with at least one comment corresponding to each of the four assessment criteria. The global comments at the end of the papers were the same in every case: they simply thanked the students for their assignments and directed them to see the marginal comments. The SREs were returned with feedback, which students then used to write their RSREs. The RSREs received two grades, each out of 100. The first grade used the same criteria laid out for the SRE, and the second grade used separate criteria related only to the improvements made from the original to the revised paper.

After the conclusion of the course, each sample was anonymized and independently analyzed by two members of the research team (the raters). The raters looked at each comment on the SRE, compared it to the corresponding passage on the RSRE, and then assessed whether and how well the student implemented the TA's feedback. The research team designed a rubric to conduct these analyses (see Figure 1) and piloted its use during a benchmarking session. The rubric required the raters to identify each comment as a Question or Statement or Imperative (Q/S/I); assign each of the 10 comments an identifier (A-J); categorize each comment as one of the four criteria (thesis, evidence, organization, or language); assess the extent of change required (low, medium, or high); assign a score to indicate the quality of the change (-1 indicates a change that degrades the paper; 0 indicates no change; 1 indicates some change, but not the full change sought; and 2 indicates the requested change); and, if appropriate, provide comments to explain the analysis. The benchmarking session confirmed that the rubric met the raters' needs and helped ensure consistency between the two sets of analysis.

The research team created a Scoring Guide to provide the raters with guidelines for completing the rubric. It defines the thesis, evidence, organization, and language criteria and offers examples of each. It also includes descriptions of what constitutes a low, medium, or high "Extent of Change Required." A "low" level of change was one that was local and simple, such as a spelling error in a single word or a paragraph that was too long and should be divided partway through. A "medium" level was a required change that was either simple but recurring, such as changing verbs to past tense throughout the paper, or complex but local, such as finding more evidence to support a specific claim. A "high" level of change was extensive and challenging, such as the scope of the entire paper being too broad and requiring a narrowing of the focus, which would, in turn, require a substantial re-write of much of the paper, or the misinterpretation of a fundamental source, which would require a good deal of extra thought and the re-writing of at least one of the paper's sections. When

Paper No.	Q/S/I	A-J	Criterion (thesis, select + present evidence, organization, language) (MC)	Criterion (thesis, select + present evidence, organization, language) (TET)	Extent of change required (low, medium, high) (MC)	Extent of change required (low, medium, high) (TET)	Score (-1,0,1,2) (MC)	Score (-1,0,1,2) (TET)	Comments (MC)	Comments (TET)
3	I	A	thesis	thesis	med	med	2	2		
		B	evidence	evidence	low	low	0	0	comment in footnote to "place publication details in parentheses"	
		C	organization	organization	low	low	2	2		
		D	organization	organization	low	low	2	2		
		E	organization	organization	low	low	2	2		
		F	language	language	low	low	2	2		
		G	organization	organization	low	low	2	2		
		H	organization	organization	low	low	2	2		
		I	organization	organization	low	low	2	2		

**Figure 2. Rubric with moderated scores**

the raters were unsure about which level of change was required by a comment, they also considered whether the implementation of feedback would take about 10 seconds, which would be "low," 10 minutes, which would be "medium," or something closer to 10 hours, which would be "high." Finally, the Scoring Guide explains and includes examples of each of the possible scores that could be assigned to the students' attempts to implement the feedback, with 0 reflecting no change, 1 reflecting partial change or an attempt to address the TA's comments, and 2 reflecting substantial improvement. Very rarely, a score of -1 was assigned to indicate that the student's change actually made the initial problem worse.

After conducting their independent analyses, the two raters met again to compare their results and generate a single set of data (see Figure 2). They went through each of the 67 sets of papers, discussed the 10 comments in each SRE and associated revisions in the RSRE, and, in cases where raters differed, reached an agreement on a single score that represented the student's implementation of each comment in the RSRE. When necessary, the raters consulted the Scoring Guide to help them arrive at a mutually agreeable number. All of the scores were reached by consensus; the raters did not average scores or use any other technique to generate a single number. At the end of this process, the raters generated 669 scores for 669 comments.<sup>1</sup> In most cases, the researchers had very similar scores in the rubric before they met to discuss their analyses; where there were minor differences, the process of discussing and coming to a consensus was relatively quick and simple.

## DATA ANALYSIS

Prior to analyzing the data using inferential statistics, descriptive statistics were used to summarize the data in frequency tables. Using PivotTables on Excel, a table was created to display the frequency distribution of data in percentages for each of the study's research questions. Generalized estimating equations (GEE) were then used to analyze the effect of the main predictor variable (i.e., type of feedback), both independently and when combined with other variables, on the dependent variable (i.e., score). The GEE method is often used when analyzing nested or hierarchical data and when the dependent variable is nominal (Hardin & Hilbe, 2003). For these reasons, GEE was an appropriate technique for this study, as the data included multiple comments nested within each essay, and the dependent variable of score consisted of three unordered categories.

To address the first research question, a GEE model with the type of feedback (question, statement, imperative) as a factor and the assignment score as a dependent variable was performed. To investigate whether the effect of feedback varies depending on the criterion (thesis, evidence, organization, language), extent of change required (low, medium, high), grade quantile (1, 2, 3), and TA (1 and 2), stratified analyses for each level of these variables with the type of feedback as a predictor of the assignment score were conducted. This strategy is an alternative to including the interaction terms between the factor variables in GEE analyses. We used this strategy to avoid overcomplication of the model with interaction terms between categorical variables. The regression coefficients for each stratified analysis were examined to detect the differences in the effect of the type of feedback on the assessment scores depending of the levels of the variables listed above.

## RESULTS

GEE proved a useful tool for considering the five research questions at the center of this study. The following section is organized around each of those questions. It describes both the results and the ways in which the GEE model generated them.

Our first and most basic research question asked the following: was a student more likely to implement feedback and, if so, to implement it well, when the feedback was phrased in the form of a question, a statement, or an imperative? To address this question using GEE, the Imperative category was used as a reference category, and the responses to the other two categories were compared to it.<sup>2</sup> As Table 1 demonstrates, there was no significant difference between Imperative and Statement feedback, but the  $p$ -value of .027 for Question feedback was significant, suggesting that students were twice less likely ( $\text{Exp}(B)=0.492$ ) to implement feedback presented in the form of a Question than they were to implement feedback presented in the form of an Imperative.

**Table 1. GEE Modelling for Research Question 1**

Parameter	B	SE	p-value	Exp(B)
Threshold				
[Score=0]	-2.169	.2593	.000	.114
[Score=1]	-.442	.2241	.048	.643
[Q/S/I=1]	-.709	.3211	.027	.492
[Q/S/I=2]	-.582	.3092	.060	.559
[Q/S/I=3]	0 <sup>a</sup>	.	.	1
(Scale)	1			

Note. Dependent Variable: Score. Model: (Threshold), Q/S/I.  
<sup>a</sup> Set to zero because this parameter is redundant..

Our second research question asked: was Question, Statement, or Imperative feedback more effective for a specific aspect of writing (i.e., the thesis, evidence, organization, or language criterion)? As Table 2 indicates, Question feedback related to the thesis criterion had a  $p$ -value of .048 which, since the beta associated with this value were negative, indicates that Question feedback was 3.3 times more likely ( $\text{Exp}(B)=0.304$ ) to result in lower scores compared to Imperative feedback. Statement feedback, on the other hand, had a  $p$ -value of .431 and therefore was not significantly different from Imperative feedback. These data suggest that Question feedback was the least likely to improve the score of the thesis criterion.

Criterion	Parameter	B	SE	$p$ -value	Exp(B)
Thesis	Threshold				
	[Score=0]	-4.026	.7584	.000	.018
	[Score=1]	.314	.3931	.424	1.369
	[Q/S/I=1]	-1.192	.6017	.048	.304
	[Q/S/I=2]	-.426	.5411	.431	.653
	[Q/S/I=3]	0 a	.	.	1
	(Scale)	1			
Evidence	Threshold				
	[Score=0]	-1.943	.3248	.000	.143
	[Score=1]	-.412	.3115	.186	.662
	[Q/S/I=1]	-.840	.3730	.024	.432
	[Q/S/I=2]	-.658	.4394	.134	.518
	[Q/S/I=3]	0 a	.	.	1
	(Scale)	1			
Organization	Threshold				
	[Score=0]	-2.305	.4037	.000	.100
	[Score=1]	-.762	.3443	.027	.467
	[Q/S/I=1]	-.902	.4636	.052	.406
	[Q/S/I=2]	-.964	.4805	.045	.381
	[Q/S/I=3]	0 a	.	.	1
	(Scale)	1			
Language	Threshold				
	[Score=0]	-1.926	.4045	.000	.146
	[Score=1]	-.623	.3167	.049	.536
	[Q/S/I=1]	-.121	.5393	.823	.886

Question feedback related to the evidence criterion had a  $p$ -value of .024 and also had a negative beta, indicating that it was 2.3 times more likely ( $\text{Exp}(B)=0.432$ ) to result in lower scores compared to Imperative feedback. Statement feedback on the evidence criterion had a  $p$ -value of .134 and so was not significant. We may conclude from this that Question feedback was again the least likely of the three feedback forms to improve the score to the evidence criterion.

Question feedback related to the organization criterion had a  $p$ -value of .052 and so was not significant. On the other hand, the  $p$ -value for Statement feedback on the organization was significant at .045, suggesting that Statement feedback was 2.6 times less likely ( $\text{Exp}(B)=0.381$ ) than Imperative feedback to improve the score on the organization criterion.

For feedback on language, there was no significant difference between the scores for Question and Imperative feedback or for Statement and Imperative feedback.

In summary, these data suggest that Question feedback was the least effective on the thesis and evidence criteria, and Statement feedback was the least effective on the organization criterion. However, for feedback on language, there was no significant difference in the effectiveness of Question or Statement feedback compared to Imperative feedback.

Our third research question asked: was a particular form of feedback more effective for a specific extent of change required? As Table 3 demonstrates, GEE modelling found no significant

difference between scores for Question and Imperative feedback or between Statement and Imperative feedback, regardless of the extent of change required. This suggests that students who received Question or Statement feedback were neither more nor less likely to implement feedback than those who receive Imperative feedback when the extent of change required was low, medium, or high.

Required	Parameter	B	SE	$p$ -value	Exp(B)
Low	Threshold				
	[Score=0]	-2.500	.2848	.000	.082
	[Score=1]	-1.227	.2356	.000	.293
	[Q/S/I=1]	-.730	.4091	.074	.482
	[Q/S/I=2]	-.626	.3596	.082	.535
	[Q/S/I=3]	0 a	.	.	1
	(Scale)	1			
Medium	Threshold				
	[Score=0]	-1.952	.3309	.000	.142
	[Score=1]	.050	.2815	.858	1.052
	[Q/S/I=1]	-.683	.3663	.062	.505
	[Q/S/I=2]	-.580	.4144	.161	.560
	[Q/S/I=3]	0 a	.	.	1
	(Scale)	1			
High	Threshold				
	[Score=0]	-2.760	.9635	.004	.063
	[Score=1]	1.111	.8653	.199	3.039
	[Q/S/I=1]	1.425	1.1193	.203	4.159
	[Q/S/I=2]	-.824	1.4801	.578	.438
	[Q/S/I=3]	0 a	.	.	1
	(Scale)	1			

The fourth research question in our study asked about the kind of student for whom different kinds of feedback were most effective, specifically: did a student's prior performance in the course correlate with the efficacy of a given form of feedback?

To conduct this analysis, we divided students into quintiles based on the grade they achieved on the SRE. Thus, the variable "SRE Grade Quintile" refers to students' prior performance in the course. Students with grades in the lowest quintile were placed into category 1; students with grades in the second-lowest quintile were placed into category 2; students with grades in the middle quintile were placed into category 3; students with grades in the second-highest quintile were placed into category 4; and students with grades in the highest quintile were placed into category 5. In other words, the first quintile contains the lowest 20% of grades, while the fifth quintile contains the top 20% of grades.

Table 4 presents the results of the GEE modelling for this research question. For students with SRE grades in the lowest and second lowest quintiles, there was no significant difference between the scores for Question and Imperative feedback or for Statement and Imperative feedback. For students with SRE grades in the third quintile, the  $p$ -value for Question feedback was significant at .047. For students in the fourth quintile, the  $p$ -value for Question feedback was also significant at .010. The betas associated with these values were negative, which suggests that Question feedback for students in the third and fourth quintiles was more likely to result in lower scores compared to Imperative feedback. There was no significant difference between the scores for Statement and Imperative feedback for this group of students. For students with SRE grades in the fifth quintile (the top 20% of grades), the  $p$ -value for Statement feedback was significant at .000. The beta associated with this value was negative, which means that Statement feedback was more likely to result in lower scores for these students compared to Imperative feedback. There was no significant difference between the scores for Question and

Imperative feedback for this group of students. To summarize these findings, it seems that for students with the lowest grades on their initial submission, the form of feedback made no significant difference to the likelihood of their implementing the feedback. For students in the middle of the grade range, feedback in either Statement or Imperative form was more likely to lead to improvement. For students in the top of the grade range, feedback in either Question or Imperative form was more likely to lead to improvement.

**Table 4. GEE Modelling for Relationship between Students' Prior Performance and Form of Feedback**

Quintile	Parameter	B	SE	p-value	Exp(B)
1	Threshold				
	[Score=0]	-1.201	.4681	.010	.301
	[Score=1]	.678	.4591	.140	1.969
	[Q/S/I=1]	-.499	.4613	.280	.607
	[Q/S/I=2]	-.563	.7157	.432	.570
	[Q/S/I=3]	0 a	.	.	1
2	Threshold				
	[Score=0]	-2.138	.3958	.000	.118
	[Score=1]	.013	.4017	.975	1.013
	[Q/S/I=1]	-.204	.6521	.754	.815
	[Q/S/I=2]	-.328	.4997	.512	.720
	[Q/S/I=3]	0 a	.	.	1
3	Threshold				
	[Score=0]	-2.277	.4039	.000	.103
	[Score=1]	-.753	.3127	.016	.471
	[Q/S/I=1]	-1.175	.5919	.047	.309
	[Q/S/I=2]	-.470	.3281	.152	.625
	(Scale)	1			

The fifth and final research question in this study asked: was a particular form of feedback more effective when it came from a specific Teaching Assistant (TA)? Both TAs in this study were senior PhD students in History with considerable teaching experience. TA 1 was male, racialized, and spoke with a British accent, which was not a local accent for the region in which the university is located. He was also the Head TA for the course. TA 2 was female, not racialized, and had an accent that would be considered typical of speakers born or raised in the region in which the university is located.

As Table 5 indicates, the *p*-value for Statement feedback from TA 1 was significant at .000. The beta associated with this value was negative, which suggests that Statement feedback was more likely to result in lower scores compared to Imperative feedback when provided by TA 1. There was no statistically significant difference between Question feedback and Imperative feedback provided by TA 1. For the second TA, there was no significant difference between the scores for Question and Imperative feedback or for Statement and Imperative feedback.

**Table 5. GEE Modelling for Relationship between TA and Form of Feedback**

TA	Parameter	B	SE	p-value	Exp(B)
TA1	Threshold				
	[Score=0]	-2.848	.2451	.000	.058
	[Score=1]	-1.088	.2205	.000	.337
	[Q/S/I=1]	-.414	.4236	.328	.661
	[Q/S/I=2]	-1.071	.2922	.000	.343
	[Q/S/I=3]	0 a	.	.	1
TA2	Threshold				
	[Score=0]	-1.600	.3573	.000	.202
	[Score=1]	.219	.3283	.504	1.245
	[Q/S/I=1]	-.366	.4145	.378	.694
	[Q/S/I=2]	-.127	.4959	.798	.881
	[Q/S/I=3]	0 a	.	.	1
(Scale)	1				

These data suggest that students were less likely to implement Statement feedback than Imperative feedback from TA 1. Otherwise, there is no significant difference in the effectiveness of Question or Statement feedback compared to Imperative feedback when it was provided by TA 2.

## IMPLICATIONS

Our study suggests that feedback in the form of a question is less effective overall than other forms of feedback (i.e., feedback phrased as statements or imperatives), especially when used to address concerns related to thesis statements or evidence. This finding might be unsurprising to some instructors, especially to those with experience teaching first-year courses where students are unaccustomed to dialogic teaching. We would note, though, that this study's results differ from the findings of other published research on the topic of feedback effectiveness and contradicts advice frequently given to instructors and TAs. Dekker et al. (2013) found that comments presented as questions were more effective than those phrased as statements. In our study, students did not seem to respond as well to feedback that positioned them as partners in a dialogue as they did to comments that made statements or made demands. This finding may be explained by studies of feedback on ELL student writing (e.g., Nurmukhamedov & Kim, 2010; Sugita, 2006) that found imperative comments most effective because ELL students preferred feedback that was more direct and required less interpretation. While we do not know the language status of the participants in this study, we do know that approximately 25% of the total student body at our institution comprise international students and that a significant number of domestic students do not speak English at home (Neebar, 2018). The students may also have found imperative comments as easier to interpret because the phrasing of such comments clearly told them what to do. Questions and statements are sometimes meant to do this as well, but indirectly, and can be misinterpreted (Hewett, 2010). Although the feedback "Can you find more evidence to support this claim?" or "More evidence could be provided to support this claim" may be meant as politely-phrased directions to offer more evidence, students could be inclined to read the feedback as, in the first instance, a genuine question (the answer to which may be, in their mind, "no, I cannot find more evidence"), and, in the second instance, a suggestion (the response to which may be, in their mind, "more evidence probably could be provided, but it doesn't have to be."). Imperatives are simply more clearly directive.

Another factor to be kept in mind is that most students in our study (81 per cent) were in their very first semester of university. They may have simply been more accustomed to directive feedback than students further along in their undergraduate careers, or they may have felt unsure of themselves as new students and thus may have responded more positively to authoritative TA guidance.

Our research also indicates that feedback in the form of a statement is most effective when identifying areas for improvement related to organization. Much of this feedback related to paragraph structure (e.g., a paragraph's length, focus, or coherence), which students may perceive as aspects of their writing governed by rules that are relatively inflexible. Thus, comments such as "this paragraph is too long" may have resonated with what students understood as a kind of truth about essay structure and

therefore may have been particularly effective when presented as a statement.

It did not seem to matter what form of feedback was employed to address matters related to language (e.g., lower-order or sentence-level concerns). In most cases, comments related to language were addressing an obvious error that required relatively little effort to correct. Many of these errors were likely the result of students' sloppiness rather than their misunderstanding of the conventions of academic writing. In these cases, once students' attention was drawn to the errors, they were probably able to recognize and fix them without needing to do additional research, reflect on the feedback, or consider different options for implementing it. There was often only one possible course of action for responding to feedback on language, irrespective of the form it took. It is important to note, however, that regardless of the feedback form, the comments always conveyed enough information for the student to make a change—the TAs never simply circled an error or used a correction symbol to note a mistake. Instead, they communicated with the student in a way that both identified the error and suggested a strategy for fixing it (e.g., "Is this the best word to use here?" or "This is not the best word to use here" or "Change this word"). Regardless of the feedback form, these instances of comments associated with language were low-hanging fruit for the student and therefore likely to be revised in the resubmitted essay.

Our study suggests that the form of feedback had no relationship with the extent of change required by the grader. Feedback advising minor or significant revisions was equally effective whether it was presented as a question, statement, or imperative. This may suggest that students were more concerned with what the feedback said than with how difficult it was to implement, or that students were uncertain of the extent of change required—especially since the same form of feedback was used throughout the paper regardless of whether the revision required was minor or significant.

This research also sought to determine whether different kinds of feedback were more or less effective for students at different levels of academic achievement in the course. For students in the lowest two quintiles, the form of feedback did not matter. This may mean that these students were simply less responsive, less motivated, or less capable of implementing feedback than other students. Students in the third and fourth quintile were less likely to implement feedback well when it was phrased as a question, while students in the top quintile were less likely to implement feedback well when it was phrased as a statement. This may suggest that only students with the strongest grades felt sufficiently confident in their skills to respond well to a dialogic approach; it may also indicate that higher-achieving students were more used to receiving and working with dialogic feedback, or that their metacognitive skills in the domain of higher education were sufficiently well-developed that they were consistently able to recognize and address a weakness in their work once it had been pointed out to them.

Finally, our research found that students were less likely to implement feedback phrased as a statement than as a question or imperative feedback from TA 1 who was male, racialized, and spoke with a British accent. We did not identify a statistically significant difference in the effectiveness of question or statement feedback compared to imperative feedback when it was provided by TA 2 who was female, not racialized, and spoke with an accent

typical of native speakers in the region. We are not certain how to interpret these results, but consider it possible that gender and race played roles in students' perception of authority.

## LIMITATIONS AND FURTHER RESEARCH

From a TA's point of view, there was a certain artificiality in the presentation of feedback required for this study. In most courses, graders adjust the form of feedback according to what they judge likely to be most effective for a given comment and a specific student, and they select the number of instances according to the needs of the assignment and the time available; in our study, by contrast, the TAs were restricted to one format and exactly ten instances per paper. They reported that using exclusively the Question form, in particular, felt inauthentic and even passive-aggressive at times, and they said that ten comments were likely insufficient for many of the papers. Furthermore, the TAs pointed out that normally they would save their lengthier or more complex feedback for the global comments at the end of a paper, but that this option was not available for the papers in this study.

As well, this study used only one assignment in the course to define prior or incoming academic ability. Grades in other courses and even, in the case of first-year students, grades from high school would provide a more detailed academic profile of the cohort that could serve as an independent variable.

Further research would clarify several findings. Interviews or focus groups could help us better understand students' perceptions of different kinds of feedback, and perhaps indicate why certain forms were more likely to lead to improvement for some issues than for others. Additional demographic data on language status might help refine the findings on whether a student's level of comfort with English influences their likelihood of responding to feedback in different forms.

## CONCLUSION

Whether phrased as a question, a statement, or an imperative, feedback has the potential to improve students' work in a revise-and-resubmit type of assignment. Generally speaking, none of the feedback was likely to lead to a decline in the quality of work. And for language errors that are fairly easy to fix, the phrasing of feedback made no significant difference. Imperative feedback had a slightly higher likelihood overall of being implemented by students, and questions had a slightly lower likelihood. Teachers should bear this in mind when determining which form an individual comment should take, and reconsider whether the frequently-recommended practice of phrasing feedback as questions is necessarily the most helpful for their students, at least in situations similar to this one (first year, with comments leading towards a resubmission). Aside from those already in the top performing group, students improved their writing more in response to statements and imperatives than in response to questions. Questions can certainly still be useful, especially when designed to get students thinking more deeply about ideas rather than conveying a judgment about what needs to be changed, but more directive feedback may often be the better choice if instructors are trying to lead students in a straightforward way to making changes in their writing.

## ACKNOWLEDGEMENTS

The authors would like to thank Andie Burazin, Rachel Courts, and Olesya Falenchuk for their assistance with the research that led to this article. We would also like to thank our colleagues at the University of Toronto Mississauga for their support, especially the faculty and staff at the Department of Historical Studies and the Robert Gillespie Academic Skills Centre. And we would like to thank our students for their patience in helping us become better teachers.

## NOTES

1. There could have been 721 scores for 721 comments, but we were unable to conduct GEE on comments that received a score of X (structured away) or -1 (the few cases where the raters agreed that the student's revision had made the essay worse). We also lost some data on two of the papers.
2. The same method (using the Imperative as a reference category and comparing the others to it) was used for all analyses.

## REFERENCES

- Bailey, R., & Garner, M. (2010). Is the feedback in higher education assessment worth the paper it is written on? Teachers' reflections on their practices. *Teaching in Higher Education*, 15 (2), 187-198.
- Bies-Hernandez, N. J. (2012). The effects of framing grades on student learning and preferences. *Teaching of Psychology* 39 (3), 176-180.
- Brown, E. & Glover, C. (2006). Evaluating written feedback. In Bryan, C. & Clegg, K. (Eds.) *Innovative Assessment in Higher Education*. (81-91), Routledge.
- Cho, K. & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction* 20 (4), 328-38.
- Davis, S. & Coleman, S. S. (2007). Re-Visiting the use of behavior therapy in graduate education: A comparative study of instructor feedback on graduate student anxiety. *Journal of Teaching in Social Work* 27 (3-4), 27-44.
- Dekker, H., Schönrock-Adema, J., Snoek, J. W., van der Molen, T., & Cohen-Schotanus, J. (2013). Which Characteristics of Written Feedback are Perceived as Stimulating Students' Reflective Competence: An Exploratory Study. *BMC Medical Education* 13 (1), 94.
- Denton, P., Madden, J., Roberts, M. & Rowe, P. (2008). Students' response to traditional and computer-assisted formative feedback: A comparative case study. *British Journal of Educational Technology* 39 (3), 486-500.
- Dlaska, A. & Krekeler, C. (2017). Does grading undermine feedback? The influence of grades on the effectiveness of corrective feedback on L2 writing. *The Language Learning Journal* 45 (2), 185-201.
- Doan, L. (2013). Is Feedback a Waste of Time? The Students' Perspective. *Journal of Perspectives in Applied Academic Practice* 1 (2), 3-10.
- Docan, T. N. (2006). Positive and negative incentives in the classroom: An analysis of grading systems and student motivation. *Journal of Scholarship of Teaching and Learning* 6 (2), 21-40.
- Donovan, P. (2014). Closing the feedback loop: physics undergraduates' use of feedback comments on laboratory coursework. *Assessment & Evaluation in Higher Education* 39 (8), 1017-1029.
- Dysthe, O, Lillejord, S., Wasson, B., and Vines, A. (2011). Productive e-feedback in higher education. Two models and some critical issues. In Ludvigsen S, Lund A, Rasmussen I, & Säljö, R (Eds.), *Learning Across Sites. New Tools, Infrastructures and Practices* (pp. 243-58). Routledge.
- Elliot, N. & Perelman, L. (2012). *Writing assessment in the 21<sup>st</sup> century: Essays in honor of Edward M. White*. Hampton Press.
- Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research* 83 (1), 70-120.
- Fisher, R., Cavanagh, J., & Bowles, A. (2011). Assisting transition to university: using assessment as a formative learning tool. *Assessment & Evaluation in Higher Education* 36 (2), 225-237.
- Forsythe, A., & Johnson, S. (2017). Thanks, but no-thanks for the feedback. *Assessment & Evaluation in Higher Education* 42 (6), 859-859.
- Freestone, N. (2009). Drafting and acting on feedback supports student learning when writing essay assignments. *Advances in Physiology Education* 33 (2), 98-102.
- Fulwiler, T. (1988). Evaluating writing across the curriculum programs. *New Directions for Teaching and Learning* 36, 61-75.
- Glover, C. & Brown, E. (2006). Written feedback for students: too much, too detailed or too incomprehensible to be effective? *Bioscience Education* 7 (1), 1-16.
- Handley, K., Price, M., and Millar, J. (2011). Beyond "doing time": Investigating the concept of student engagement with feedback. *Oxford Review of Education* 37 (4), 543-60.
- Hardin, J.W. and Hilbe, J.M. (2003). *Generalized Estimating Equations*. Chapman & Hall.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77 (1), 81-113.
- Hepplestone, S., & Chikwa, G. (2014). Understanding how students process and use feedback to support their learning. *Practitioner Research in Higher Education* 8 (1), 41-53.
- Hewett, B.L. (2010). *Online Writing Conference: A guide for teachers and tutors*. Heinemann.
- Hounsell, D. (2007). Toward more sustainable feedback to students. In Boud, D. & Falchikov, N. (Eds.), *Rethinking Assessment in Higher Education: Learning for the Longer Term*. (pp. 101-113). Routledge.
- Huxham, M. (2007). Fast and effective feedback: Are model answers the answer? *Assessment & Evaluation in Higher Education* 32 (6), 601-11.
- Hyatt, D. (2005). "Yes, a very good point!": A critical genre analysis of a corpus of feedback commentaries on master of education assignments. *Teaching in Higher Education* 10 (3), 339-53.
- Hyland, P. (2000). Learning from feedback on assessment. In Booth, A. & Hyland, P. (Eds.), *The Practice of University History Teaching*. (pp. 233-247). Manchester University Press.
- Jonsson, A. (2012). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education* 14 (1), 63-76.
- Kaler, M., & Evans-Tokaryk, T. (2019). Reflecting on Assessment: Strategies and Tools for Measuring the Impact of a Canadian WAC Program. *Discourse and Writing/Rédactologie* 29, 107-132.
- Kluger, A. N., and A. DeNisi, A. (1996). The effects of feedback interventions on performance: Historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 119 (2), 254-284.



- Lipnevich, A., & Smith, J. (2009). Effects of Differential Feedback on Students' Examination Performance. *Journal of Experimental Psychology* 15 (4), 319–333.
- Mulliner, E., & Tucker, M. (2017). Feedback on feedback practice: perceptions of students and academics. *Assessment & Evaluation in Higher Education* 42 (2), 266-288.
- Mutch, A. (2003). Exploring the practice of feedback to students. *Active Learning in Higher Education* 4 (1), 24–38.
- Neebar, L. (2018). SEM November Update November 21, 2018 as presented to Campus Council. University of Toronto Mississauga.
- Nicol, D.J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education* 31 (2), 199-218.
- Nurmukhamedov, U. & Kim, S. H. (2010). "Would You Perhaps Consider...": Hedged Comments in ESL Writing. *ELT Journal* 64 (3), 274-282.
- O'Neill, P., Moore, C., & Huot, B. (2009). *A Guide to college writing assessment*. University Press of Colorado.
- Pruchnik, J. et al. (2018). Slouching towards sustainability: Mixed methods in the assessment of student writing. *Journal of Writing Assessment* 11 (1)
- Riddell, J. (2015). Performance, feedback, and revision: metacognitive approaches to undergraduate essay writing. *Collected Essays on Learning and Teaching* 8, 79-95.
- Robinson, S., Pope, D., & Holyoak, L. (2013). Can we meet their expectations? Experiences and perceptions of feedback in first year undergraduate students. *Assessment & Evaluation in Higher Education* 38 (3), 260–272.
- Ryan, T., & Henderson, H. Feeling feedback: students' emotional responses to educator feedback. *Assessment & Evaluation in Higher Education* 43 (6), 880-892.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* 18 (2), 119-144.
- Skinner, K. (2014). Bridging gaps and jumping through hoops: First-year History students' expectations and perceptions of assessment and feedback in a research-intensive UK university. *Arts & Humanities in Higher Education* 13 (4), 359–376.
- Stern, L. A. & Solomon, A. (2006). Effective faculty feedback: The road less traveled. *Assessing Writing* 11 (1), 22-41.
- Sugita, Y. (2006). The impact of teachers' comment types on students' revision. *ELT Journal* 60 (1), 34-41.
- Tang, J., & Harrison, C. (2011). Investigating university tutor perceptions of assessment feedback: three types of tutor beliefs. *Assessment & Evaluation in Higher Education* 36 (5), 583–604.
- Tippin, G. K., Lafreniere, K. D., and Page, S. (2012). Student perception of academic grading: Personality, academic orientation, and effort. *Active Learning in Higher Education* 13 (1), 51-61.
- Torres, J.T., & Anguiano, C. J. (2016). Interpreting feedback: A discourse analysis of teacher feedback and student identity. *Practitioner Research in Higher Education Journal* 10 (2), 2-11.
- Vardi, I. (2013). Effectively feeding forward from one written assessment task to the next. *Assessment & Evaluation in Higher Education* 38 (5), 599–610.
- Walker, M. (2009). An investigation into written comments on assignments: Do students find them usable? *Assessment & Evaluation in Higher Education* 34 (1), 67–78.
- Winstone, N. E., Nash, R.A., Rountree, J., & Parker, M. (2017) "It'd be useful, but I wouldn't use it": Barriers to university students' feedback seeking and recipience. *Studies in Higher Education* 42 (11), 2026-2041.
- Yancey, K. B. & Huot, B. (Eds.) (1997). *Assessing writing across the curriculum: Diverse approaches and practices*. Ablex.