Feedback That Leads to Improvement in Student Essays: Testing the Hypothesis that "Where to Next" Feedback is Most Powerful

John Hattie[1,2], Jill Crivelli[3], Kristin Van Gompel[3], Patti West-Smith[3]* and Kathryn Wike[3]

[1]Graduate School of Education, University of Melbourne, Melbourne, VI, Australia, [2]Hattie Family Foundation, Melbourne, VI, Australia, [3]Turnitin, LLC, Oakland, CA, United States

Feedback is powerful but variable. This study investigates which forms of feedback are more predictive of improvement to students' essays, using *Turnitin Feedback Studio*–a computer augmented system to capture teacher and computer-generated feedback comments. The study used a sample of 3,204 high school and university students who submitted their essays, received feedback comments, and then resubmitted for final grading. The major finding was the importance of "where to next" feedback which led to the greatest gains from the first to the final submission. There is support for the worthwhileness of computer moderated feedback systems that include both teacher- and computer-generated feedback.

Keywords: feedback, essay scoring, formative evaluation, summative evaluation, computer-generated scoring, instructional practice, instructional technologies, writing

INTRODUCTION

One of the more powerful influences on achievement, prosocial development, and personal interactions is feedback–but it is also remarkably variable. Kluger and DeNis (1996) completed an influential meta-analysis of 131 studies and found an overall effect on 0.41 of feedback on performance and close to 40% of effects were negative. Since their paper there have been at least 23 meta-analyses on the effects of feedback, and recently Wisniewski et al. (2020) located 553 studies from these meta-analyses ($N = 59{,}287$) and found an overall effect of 0.53. They found that feedback is more effective for cognitive and physical outcome measures than for motivational and behavioral outcomes. Feedback is more effective the more information it contains, and praise (for example), not only includes little information about the task, but it can also be diluting as receivers tend to recall the praise more than the content of the feedback. This study investigates which forms of feedback are more predictive of improvement to students' essays, using Turnitin Feedback Studio–a computer augmented system to capture teacher- and computer-generated feedback comments.

Hattie and Timperley (2007) defined feedback as relating to actions or information provided by an agent (e.g., teacher, peer, book, parent, internet, experience) that provides information regarding aspects of one's performance or understanding. This concept of feedback relates to its power to "fill the gap between what is understood and what is aimed to be understood" (Sadler, 1989). Feedback can lead to increased effort, motivation, or engagement to reduce the discrepancy between the current status and the goal; it can lead to alternative strategies to understand the material; it can confirm for the student that they are correct or incorrect, or how far they have reached the goal; it can

indicate that more information is available or needed; it can point to directions that the students could pursue; and, finally, it can lead to restructuring understandings.

To begin to unravel the moderator effects that lead to the marked variability of feedback, Hattie and Timperley (2007) argued that feedback can have different perspectives: "feed-up" (comparison of the actual status with a target status), "feed-back" (comparison of the actual status with a previous status), and "feed-forward" (explanation of the target status based on the actual status). They claimed that these related to the three feedback questions: Where am I going? How am I going? and Where to next? Additionally, feedback can be differentiated according to its level of cognitive complexity: It can refer to a task, a process, one's self-regulation, or one's self. Task level feedback means that someone receives feedback about the content, facts, or surface information (How well have the tasks been completed and understood?). Feedback at the level of process means that a person receives feedback on the processes or strategies of his or her performance (What needs to be done to understand and master the tasks?). Feedback at the level of self-regulation means that someone receives feedback about the individual's regulation of the strategies they are using to their performance (What can be done to manage, guide, and monitor your own way of action?). The self-level focuses on the personal characteristics of the feedback recipient (often praise about the person). One of the arguments about the variability is that feedback needs to focus on the appropriate question and the optimal level of cognitive complexity. If not, the message can easily be ignored, misunderstood, and of low value to the recipient.

Another important distinction is between the giving and receiving of feedback. Students are more often the receiver, and this is becoming more a focus of research. Students indicate a preference for feedback that is specific, useful, and timely (Pajares and Graham, 1998; Gamlem and Smith, 2013), relative to the criteria or standards they are assessed against (Brown, 2009; Beaumont et al., 2011), and do not mind what form it comes provided they see it as informative to improve their learning. Dawson et al. (2019) asked teachers and students about what leads to the most effective feedback. The majority of teachers argued it was the design of the task that lead to better feedback and students argued it was the quality of the feedback provided to them in teacher comments that led to improvements in performance.

Brooks et al. (2019) investigated the prevalence of feedback relative to these three questions in upper elementary classrooms. They recorded and transcribed 12 h of classroom audio based on 1,125 grade five students from 13 primary schools in Queensland. The researchers designed a questionnaire to measure the usefulness of feedback aligned with the three feedback questions ("Where am I going?" "How am I going?" "Where to next?") along with three of the four feedback levels (task, process, and self-regulation). Results indicated that of the three feedback questions, "How am I going?" (Feed-back) was by far the most prominent, accounting for 50% of total feedback words. This was followed by "Where am I going?" (Feed-up) (31%) and "Where to next?" (Feed-forward) (19%). When considering the

focus of verbal feedback, 79% of the feedback was at the task level, 16% at process level, and <1% at the self level. The findings of such studies are significant in relation to the gap between literature and practice, which indicates that we need to know more about how effective feedback interventions are enacted in the classroom.

Mandouit (2020) developed a series of feedback questions from an intensive study of student conceptions of feedback. He found that students sought feedback as to how to "elaborate on ideas" and "how to improve." They wanted feedback that would not only help them "next time" they complete a similar task in the future, but that would help them develop the ability to think critically and self-regulate moving forward. It is these transferable skills and understandings that students consider as important, but, as identified in this study, challenged teachers in practice as it was rarely offered. His student feedback model included four questions: Where have I done well? Where can I improve? How do I improve? What do I do next time?

One often suggested method of improving the nature of feedback is to administer it via computer-based systems. Earlier synthesis of this literature tended to focus on task or item-specific level and investigating the differences between knowledge of results (KR), knowledge of correct response (KCR), and elaborated feedback (EF). Van der Kleij, Feskens, and Eggen (2015), for example, used 70 effects from 40 studies of item-based feedback in a computer-based environment on students' learning outcomes. They showed that elaborated feedback (e.g., providing an explanation) produced larger effect-sizes (EF = 0.49) than feedback regarding the correctness of the answer (KR = 0.05) or providing the correct answer (KCR = 0.32). Azevedo and Bernard (1995) used 22 studies on the effects of feedback on learning from computer-based instruction with an overall effect of 0.80. Immediate feedback had an effect of 0.80 and delayed 0.35, but they did not relate their findings to specific feedback characteristics. Jaehnig and Miller (2007) used 33 studies and found elaborated feedback was more effective than KCR, and KCR was more effective than KR. The major message is the computer-delivered elaborated feedback has the largest effects.

The Turnitin Feedback Studio Model: Background and Existing Research

Turnitin Feedback Studio, one such computer-based system, is most known for its similarity checking, powered by a comprehensive database of academic, internet, and student content. Beyond that capability, however, Feedback Studio also offers functionality to support both effective and efficient options for grading and, most relevant to this study, providing feedback. Inside the system, the Feedback Studio model allows for multiple streams of feedback, depending on how instructors opt to utilize the system, with both automated options and teacher-generated options. The primary automated option is for grammar feedback, which automatically detects issues and provides guidance through an integration with the e-rater® engine from ETS (https://www.ets.org/erater). Even this option allows for customization and additional guidance, as instructors are able

to add elaborative comments to the automated feedback. Outside of the grammar feedback, the remaining capabilities are manual, in that instructors identify the instances requiring feedback and supply the specific feedback content. Within this structure, there are still multiple avenues for providing feedback, including inline comments, summary text or voice comments, and Turnitin's trademarked QuickMarks®. In each case, instructors determine what student content requires commenting and then develop the substance of the feedback.

As a vehicle for providing feedback on student writing, Turnitin Feedback Studio offers an environment in which the impact of feedback can be leveraged. Student perceptions about the kinds of feedback that most impact their learning align to findings from scholarly research (Kluger and DeNis, 1996; Wisniewski et al., 2020). Periodically, Turnitin surveys students to gauge different aspects of the product. In studies conducted by Turnitin, student perceptions of feedback over time fall into similar patterns as in outside research. For example, a 2013 survey about students' perceptions of the value, type, and timing of instructor feedback reported that 67% of students claimed receiving general, overall comments, but only 46% of those students rated the general comments as "very helpful." Respondents from the same study rated feedback on thesis/development as the most valuable, but reported receiving more feedback on grammar/mechanics and composition/structure (Turnitin, 2013). Turnitin (2013) suggests the disconnect between the receipt of general, overall comments compared to the perceived value provides further support that students value more specific feedback, such as comments on thesis/development.

Later, an exploratory survey examining over 2,000 students' perceptions on instructor feedback asked students to rank the effectiveness of types of feedback. The survey found that the greatest percentage (76%) of students reported suggestions for improvement as "very" or "extremely effective." Students also highly perceived feedback such as specific notes written in the margins (73%), use of examples (69%), and pointing out mistakes as effective (68%) (Turnitin, 2014). Turnitin (2014) proposes, "The fact that the largest number of students consider suggestions for improvement to be "very" or "extremely effective" lends additional support to this assertion and also strongly suggests that students are looking at the feedback they receive as an extension of course or classroom instruction."

Turnitin found similar results in a subsequent survey that asked students about the helpfulness of types of feedback. Students most strongly reported suggestions for improvement (83%) as helpful. Students also preferred specific notes (81%), identifying mistakes (74%), and use of examples (73%) as types of feedback. Meanwhile, the least helpful types of feedback reported by students were general comments (38%) and praise or discouragement (39%) (Turnitin, 2015). As a result of this survey data, Turnitin (2015) proposed that "Students find specific feedback most helpful, incorporating suggestions for improvement and examples of what was done correctly or incorrectly." The same 2015 survey found that students consider instructor feedback to be just as critical for their learning as doing homework, studying, and listening to

lectures. From the 1,155 responses, a majority of students (78%) reported that receiving and using teacher feedback is "very" or "extremely important" for learning. Turnitin (2015) suggests that the results from the survey demonstrates that students consider feedback to be just as important to other core educational activities.

Turnitin's own studies are not the only evidence of these trends in students' perceptions of feedback. In a case study examining the effects of Turnitin's products on writing in a multilingual language class, Sujee et al. (2015) found that the majority of the learners expressed that Turnitin's personalized feedback and identification of errors met their learning needs. Students appreciated the individualized feedback and claimed a deeper engagement with the content. Students were also able to integrate language rules from the QuickMark drag-and-drop comments, further strengthening the applicability in a second language classroom (Sujee et al., 2015). A 2015 study on perceptions of Turnitin's online grading features reported that business students favored the level of personalization, timeliness, accessibility, and quantity and quality of receiving feedback in an electronic format (Carruthers et al., 2015). Similarly, a 2014 study exploring the perceptions of healthcare students found that Turnitin's online grading features enhanced timeliness and accessibility of feedback. In particular regard to the instructor feedback tools in Turnitin Feedback Studio (collectively referred to as GradeMark), students valued feedback that was more specific since instructors could add annotated comments next to students' text. Students claimed it increased meaningfulness of feedback which further supports the GradeMark tools as a vehicle for instructors to provide quality feedback (Watkins et al., 2014). In both studies, students expressed interest in using the online grading features more widely across other courses in their studies (Watkins et al., 2014; Carruthers et al., 2015).

In addition to providing insight about students' perception of what is most effective, Turnitin studies also surfaced issues that students sometimes encounter with feedback provided inside the system. Part of the 2015 study focused on how much students read, use, and understand feedback they receive. Turnitin (2015) reports that students most often read a higher percentage of feedback than they understand or apply. When asked about barriers to understanding feedback, students who claimed to understand a minimal amount of instructor feedback (13%) reported that most often/always the largest challenges were: comments had unclear connections to the student work or assignment goals (44.8%), feedback was too general (42.6%), and they received too many comments (31.8%) (Turnitin, 2015). Receiving feedback that was too general was also considered a strong barrier for students who claimed to understand a moderate or large amount of feedback.

Research Questions

From studies investigating students' conceptions of feedback, Mandouit (2020) found that while they appreciated feedback about "where they are going", and "how they are going", they saw

feedback mainly in terms of helping them know where to go next in light of submitted work. Such "where to next" feedback was more likely to be enacted.

This study investigates a range of feedback forms, and in particular investigates the hypothesized claim that feedback that leads to "where to next" decisions and actions by students is most likely to enhance their performance. It uses Turnitin Feedback Studio to ask about the relation of various agents of feedback (teacher, machine program), and codes the feedback responses to identify which kinds of feedback are related to the growth and achievement from first to final submission of essays.

METHOD

Sample

In order to examine the feedback that instructors have provided on student work, original student submissions and revision submissions, along with corresponding teacher- and machine intelligence-assigned feedback from Feedback Studio were compiled by the Turnitin team. All papers in the dataset were randomly selected using a postgreSQL random () function. A query was built around the initial criteria to fetch assignments and their associated rubrics. The initial criteria included the following: pairs of student original drafts and revision assignments where each instructor and each student was a member of one and only one pairing of assignments; assignments were chosen without date restrictions through random selection until the sample size (<3,000) had been satisfied; assignments were from both higher education and secondary education students; assignment pairs where the same rubric had been applied to both the original submission and the revision submission and students had received scores based on that rubric; any submissions with voice-recorded comments were excluded; and submissions and all feedback were written only in the English language. Throughout the data collection process, active measures were taken to exclude all personally identifiable information, including student name, school name, instructor name, and paper content, in accordance with Turnitin's policies. The Chief Security Officer of Turnitin conducted a review of this approach prior to completion. After the dataset was returned, an additional column was added that assigned a random number to each data item. That random number column was then sorted and returned the final dataset of student submissions and resubmissions in random order, from which the final sample of student papers were identified for analysis.

The categories for investigation included country of student, higher education or high school setting, number of times the assignment was submitted, date and time of submission, details regarding the scoring of the assignment (like score, possible points, and scoring method), and details regarding feedback that was provided on the assignment (like mark type, page location of each mark, title of each mark, and comment text associated with each mark), and two

outcome measures–achievement and growth from time 1 to time 2.

There were 3,204 students who submitted essays for feedback on at least two occasions. About half (56%) were from higher education and the other half (44%) from secondary schools. The majority (90%) were from the United States, and the others were from Australia (5.2%), Japan (1.5%), Korea (0.8%), India (0.5%), Egypt (0.5%), the Netherlands (0.4%), China (0.4%), Germany (0.3%), Chile (0.2%), Ecuador (0.2%), Philippines (0.2), and South Africa (0.03%). Within the United States, students spanned 13 states, with the majority coming from California (464), Texas (412), Illinois (401), New York (256), New Jersey (193), Washington (93), Wisconsin (91), Missouri (81), Colorado (67), and Kentucky (61).

Procedures

In this study, pairs of student-submitted work—original drafts and revisions of those same assignments—along with the feedback that was added to each assignment, were examined. Student assignments were submitted to the Turnitin Feedback Studio system as part of real courses to which students submit their work via online, course-specific assignment inboxes. Upon submission, student work is reviewed by Turnitin's machine intelligence for similarity to other published works on the Internet, submissions by other students, or additional content available within Turnitin's extensive database. At this point in the process, instructors also have the opportunity to provide feedback and score student work with a rubric.

Feedback streams for student submissions in Turnitin Feedback Studio are multifaceted. At the highest level, holistic feedback can be provided in the Feedback Summary panel as a text comment. However, if instructors wish to embed feedback directly within student submissions, there are several options. First, the most prolific feature of Turnitin Feedback Studio is QuickMarks™, a set of reusable drag-and-drop comments derived from corresponding rubrics aligned to genre and skill-level criteria. Instructors may also choose to create their own QuickMarks and rubrics to save and reuse on future submissions. When instructors wish to craft personalized feedback not intended for reuse, they may leave a bubble comment, which appears in a similar manner to the reusable QuickMarks, or an inline comment that appears as a free-form text box they can place anywhere on the submission. Instructors also have access to a strikethrough tool to suggest that a student should delete the selected text. Automated grammar feedback can be enabled as an additional layer, offering the identification of grammar, usage, mechanics, style, and spelling errors. Instructors have the option to add an elaborative comment, including hyperlinks to instructional resources, to the automated grammar and mechanics feedback (delivered via e-rater®) and Turnitin QuickMarks. Finally, rubrics and grading tools are available to the teacher to complete the feedback and scoring process.

Within the prepared dataset, paired student assignments were presented for analysis. Work from each individual student was
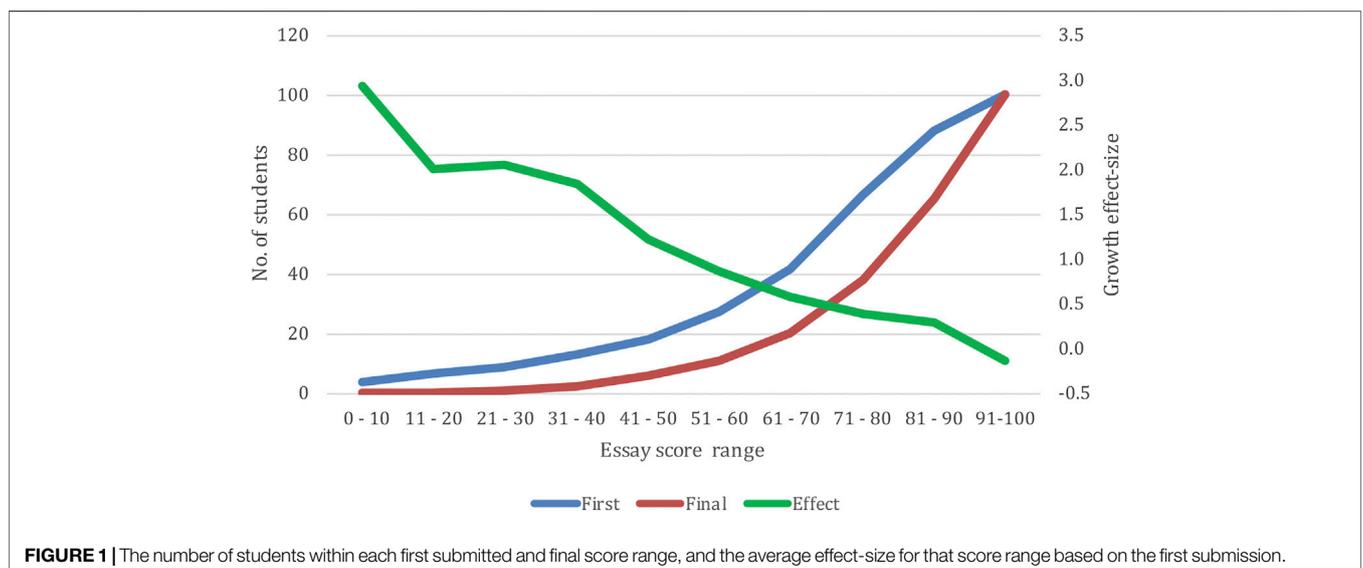
TABLE 1 | Codes and description of attributes coded for each essay.

Code	Description
General comment	This is the general comment(s) from the marker (usually at end) of section, that suggest where the student could modify to add. (e.g., "the building blocks of your paper are good, but you need to use them to further your thesis. The problem is that your thesis is unclear at this point".
Where to next? General	General comment that identifies ways to improve and specific about where to next to the student (in general comments) (e.g., "your topic score is very easy to fix. Just focus on the money and financial topics in the paper. Thesis: State your argument in the intro.. Also use a form of in text citation that is more compact.")
Where to next? Specific	Specific "where to next" comments for a student throughout the essay (e.g., "you need to cite corsetti for the ideas in this paragraph. The "two main interpretations" are right out of that paper")
Praise	# Of praise comments such as "good effort with … " and "excellent use of … "
Probes	# Of questions asking for more clarification (e.g., "I'd also suggest that you think a little bit more about Lucile's act or resistance? Is it passive-aggressive? Or is it just passive?")
Needs support	# Of statements asking for more supporting clarification (e.g., to elicit student thinking and decision-making related to their learning)
Grammar	# Of statements about grammar, spelling, or expression issues
Word count	Coded yes (1) if there was a comment relating to the word count (e.g., 'this essay is too short', the teacher stated that they stopped reading or marking at a certain point)
References	# Of comments about references (e.g., comments of student plagiarism, formatting, and not following the appropriate manual of style)
Seek additional help	# Of comments referring the student to seek additional help (e.g., 'make an appointment with the writing center,' 'come see me to work on this, or 'look up (specific) referencing manual'
Uncodeable symbols	# Of uncodeable symbols such as "????", ")"
Unclear comments	# Of unclear comments, such as "so?"/"delete" and were not only unclear to the coders but probably to the students also
Total no. of comments	A count of the number of unique comments (some may be lengthy) provided on the essay

used only once, but appeared as a pair of assignments, comprising an original, "first draft" submission, and then a later "revision" submission of the same assignment by the same student. The first set of feedback thus can be considered formative, and the latter summative feedback. For each pair of assignments, the following information was reported: institution type, country, and state or province for each individual student's work. Then, for both the original assignment submission and the revision assignment submission, the following information was reported: assignment ID, submission ID, number of times the assignment was submitted, date and time of submission, details regarding the scoring of the assignment (like score, possible points, and scoring method), and details regarding feedback that was provided on the assignment (like mark type, page location of each mark, title of each mark, and comment text associated with each mark). Prior to the analysis, definitions of all terms included within the dataset were created collaboratively and recorded in a glossary to ensure a common understanding of the vocabulary.

Some of the essays had various criteria scores (such as ideas, organization, evidence, style), but in this study only the total score was used. The assignments were marked out of differing totals so

FIGURE 1 | The number of students within each first submitted and final score range, and the average effect-size for that score range based on the first submission.

all were converted to percentages. On average, there were 19 days between submissions (SD = 18.4). Markers were invited by the Turnitin Feedback Studio processes to add comments to the essays and these were independently coded into various categories (see **Table 1**). One researcher was trained in applying the coding manual, and close checking was undertaken for the first 300 responses, leading to an inter-rater reliability in excess of 0.90, with all disagreements negotiated.

There were two outcome measures. The first is the final score after the second submission, and the growth effect-size between the score after the first submission (where the feedback was provided) and the final score. The effect-size for each student was calculated using the formula for correlated or dependent samples.

$$(\text{Post-test} - \text{Pre-test score}) / \text{SQRT}\left(\text{var}_{pre} + \text{var}_{post}\right.$$
$$\left. - 2 * r * \text{SD}_{pre} * \text{SD}_{post}\right)$$

A structural model was used to relate the feedback types with the final and growth effect-size. A multivariate analysis of variance investigates the nature of changes in means from the first to final scores, moderated by level of schooling (secondary, university). A regression was used to identify the source of feedback relative to the growth and final scores.

RESULTS

The average score at Time 1 was 71.34 (SD = 19.91) and at Time 2 was 82.97 (SD = 15.03). The overall effect-size was 0.70 (SD = 0.97) with a range from −2.26 to 4.97. The correlation between Time 1 and 2 scores was 0.60.

Figure 1 shows the number of students in each score range, and the average effect-size for that score range. Not surprising, the opportunity to improve (via the effect-size) is greater for those who scored lower in their essays at Time 1. There were between 1 and 139 total comments for the first submission essays with an average of 14 comments per essay (**Table 2**). The most common comments related to Where to next–Specific (5.9), Needs support (4.5), Where to next–General (3.8), and Probes (2.3). The next set of common comments were about style such as references (2.0), Unclear comments (1.9),

TABLE 2 | Range, mean, and standard deviation of feedback comments for first submission essay.

Feedback forms	Range	Mean	SD
General comment	0–1	0.97	0.16
Where to next? General	0–66	3.77	3.10
Where to next? Specific	0–47	5.89	6.37
Praise	0–21	1.25	1.88
Probes	0–30	2.31	3.24
Needs support	0–27	4.51	3.87
Grammar, punctuation, spelling	0–35	1.73	2.68
Word count	0–9	0.10	0.39
References	0–30	2.02	2.79
Seek additional help	0–6	0.22	0.56
Uncodeable symbols	0–12	0.15	0.71
Unclear comments	0–81	1.89	4.06
Total no. of comments	0–139	14.58	11.06

Grammar, punctuation, and spelling (1.7). There was about 1 praise comment per essay, and the other forms of feedback were more rare (Seek additional help (0.22), Uncodeable symbols (0.15), and Word count (0.10). The general message is that instructors were mostly focused on improvement, then on the style aspects of the essays.

There are two related dependent variables–the relation between the comments and the Time 2 grade, and to the improvement between Time 1 and Time 2 (the growth effect-size). Clearly, there is a correlation between Time 2 and the effect-size (as can be seen in **Figure 1**) but it is sufficiently low ($r = 0.19$) to warrant asking about the differential relations of the comments to these two outcomes.

A covariance analysis using SEM (Amos, Arbuckle, 2011) identified the statistically significant correlates of the Time 2 and growth effect-sizes. Using only these forms of feedback statistically significant, then a reduced model was run to optimally identify the weights of the best sub-set. The reduced model (chi-square = 18,466, df = 52) was statistically significantly better fit (chi-square = 19,686, df = 79; Δchi-square = 1,419, df = 27, $p <. 001$).

Thus, the best predictors of the growth improvement from Time 1 to Time 2 were the number of comments (the more comments given, the more likely the essay improved), and Specific and General Where to next comments (**Table 3**). The best predictors of the overall Time 2 performance were Praise; and the comments that led to the lowest improvement included Praise, Probes, Grammar, Referencing, and Unclear comments. It is worth noting that Praise for a summative outcome is positive, but for formative is negative.

A closer investigation was undertaken to see if Praise indeed has a dilution effect. Each student's first submission was coded as having no Praise and no Where-to-next ($N = 334$), only Praise ($N = 416$), only Where-to-next ($N = 1,113$), and Praise and Where-to-next feedback ($N = 1,434$). When the first two sets were considered, the improvement was appreciably lower where there was Praise compared to no Praise and no Where-to-next (Mn = −0.21 vs. 0.40), and similar compared to Where-to-next and "Praise and Where-to-next" (Mn = 0.89 vs. 0.89).

There was an overall mean difference in the Time 1, Time 2, and growth effect-size relating to whether the student was at University or within a High School (Wilks Lambda = 0.965, Mult. F = 57.68, df = 2, 3,189, $p < 0.001$; **Table 4**). There were no differences between the mean scores at Time 1, but the University students made the greatest growth between Time 1 and Time 2, and thence in the final Time 2 grade. There were more comments for University students inviting students to seek additional help, and more Where to next comments. The instructors of University students gave more specific and general Where to next feedback comments (4.11, 6.55 vs. 3.30, 4.87) than did the instructors/markers of the secondary students. There were no differences in the number of words in the comments, Praise, the provision of general comments or not, uncodeable comments, and referencing.

For University students, the highest correlates of the specific coded essay comments included Where to next, the number of comments, General and Specific Where to next, Need support, Seek additional help, the total number of comments, and negatively related to Praise (**Table 5**). For secondary students, the highest correlates were Where to next, Need support, and negative to Praise.

TABLE 3 | Standardized structural weights for the full and reduced covariance analyses for the feedback forms.

	Full covariance model			Reduced model		
	Time 2	p	Growth	p	Time 2	Growth
General comment	−0.01	0.485	−0.04	0.004		
Where to next?–General	−0.02	0.233	0.13	<0.001		0.117
Where to next? - Specific	−0.07	<0.001	0.13	<0.001	−0.041	0.112
Praise	0.22	<0.001	−0.15	<0.001	0.231	−0.158
Probes	−0.09	<0.001	−0.15	<0.001	−0.096	−0.149
Needs support	−0.08	<0.001	−0.04	0.004	−0.089	
Grammar punctuation spelling	0.02	0.226	−0.20	<0.001		−0.188
Wordcount	−0.08	<0.001	−0.02	0.278	−0.082	
References formatting	−0.02	0.258	−0.20	<0.001		−0.193
Seek additional help	0.02	0.185	0.04	0.012		
Uncodeable	0.01	0.477	0.03	0.014		
Unclear	0.07	<0.001	−0.16	<0.001	0.087	−0.144
No. comments	0.03	0.051	0.44	<0.001		0.432

TABLE 4 | Means, standard deviations, effect-sizes, and analysis of variance statistics of comparisons between University and Secondary students.

	University	SD	Secondary	SD	Effect-size	F	df	p
Time 1	71.20	20.18	71.52	19.56	−0.02	0.20	1, 3,190	0.65
Time 2	84.88	14.06	80.52	15.87	0.29	67.53	1, 3,190	<0.001
Effect-size	0.82	1.07	0.54	0.81	0.30	67.41	1, 3,190	<0.001
No	1792		1,400					

TABLE 5 | Correlations between the forms of feedback for the university and secondary students.

	University	p	Secondary	p
General comment (Y/N)	−0.02	0.356	−0.01	0.655
Where to next? - General	0.22	<0.001	0.14	<0.001
No. comments	0.15	<0.001	0.07	0.008
Where to next? - Specific	0.14	<0.001	0.04	0.159
Praise	−0.12	<0.001	−0.12	<0.001
Probes	0.10	<0.001	0.08	0.004
Needs support	0.23	<0.001	0.14	<0.001
Grammar punctuation spelling	−0.07	0.003	0.01	0.91
Word count	0.05	0.037	−0.01	0.815
References formatting	0.02	0.405	−0.08	0.004
Seek additional help	0.12	<0.001	−0.03	0.202
Uncodeable	0.09	<0.001	0.01	0.793
Unclear	−0.02	0.326	0.03	0.298
Number	−0.08	0.001	−0.08	0.004
Total no	0.14	<0.001	0.07	0.011

There are five major forms of feedback provisions, and the most commonly used were e-rater[®] (grammar), QuickMarks (drag-and-drop comments), and teacher-provided comments. There were relatively few inline (instructor brief comments), and strikethroughs (**Table 6**). Across all essays, there were significant relations between teacher inline, QuickMarks, and strikethroughs with the growth impact over time. Perhaps not surprising, these same three correlated negatively with the performance at first submission as these had the greatest opportunity for teacher comments.

CONCLUSION

Feedback can be powerful but it is also most variable. Understanding this variability is critical for instructors who aim to improve their students' proficiencies. There is so much advice about feedback

TABLE 6 | Means, standard deviations, and correlations between forms of feedback provision and first submission, final submission, and growth effect-sizes.

	Mean	SD	% Comments	First	Final	Effect
Comments	3.37	8.08	27	−0.02	−0.02	0.06
E-rater	6.76	12.16	34	0.02	−0.08	−0.10
Inline	0.60	2.98	4	−0.14	0.08	0.27
QuickMark	3.92	8.90	30	−0.19	−0.02	0.28
Strikethrough	0.54	3.18	5	−0.17	0.05	0.26
Total # comments	28.34	26.62		−0.13	0.01	0.19

sandwiches (including a positive comment, then specific feedback comment, then another positive comment), increasing the amount of feedback, the use of praise about effort, and debates about grades or comments, but these all ignore the more important issue about how any feedback is heard, understood, and actioned by students. There is also a proliferation of computer-aided tools to improve the giving of feedback, and with the inclusion of artificial intelligence engines, these are proffered as solutions to also reduce the time and investment by instructors in providing feedback. The question addressed in this study is whether the various forms of feedback is "heard and used by students" leading to improved performance.

As Mandouit (2020) argued, students prefer feedback that assists them to know where to learn next, and then how to attain this "where to next" status; although this appears to be a least frequent form of feedback (Brooks et al., 2019). Others have found that more elaborate feedback produces greater gains in learning than feedback about the correctness of the answer, and this is even more likely to be the case when asked for essays rather than closed forms of answering (e.g., multiple choice).

The major finding was the importance of "where to next" feedback, which lead to the greatest gains from the first to the final submission. No matter whether more general or quite specific, this form of feedback seemed to be heard and actioned by the students. Other forms of feedback helped, but not to the same magnitude; although it is noted that the quantity of feedback (regardless of form) was of value to improve the essay over time.

Care is needed, however, as this "where to next" feedback may need to be scaffolded on feedback about "where they are going" and "how they are going," and it is notable that these students were not provided with exemplars, worked examples, or scoring rubrics that may change the power of various forms of feedback, and indeed may reduce the power of more general forms of "where to next" feedback.

In most essays, teachers provided some praise feedback, and this had a negative effect on improvement, but a positive effect on the final submission. Praise involves a positive evaluation of a student's person or effort, a positive commendation of worth, or an expression of approval or admiration. Students claim they like praise (Lipnevich, 2007), and it is often claimed praise is reinforcing such that it can increase the incidence of the praise behaviors and actions. In an early meta-analysis, however, Deci et al. (1999) showed that in all cases, the effects of praise were negative on increasing the desired behavior; task noncontingent–praise given from something other than engaging in the target activity (e.g., simply participating in the lesson) ($d = -0.14$); task contingent–praise given for doing or completing the target activity ($d = -0.39$); completion contingent–praise given specifically for performing the activity well, matching some standard of excellence, or surpassing some specific criterion ($d = -0.44$); engagement contingent–praise dependent on engaging in the activity but not necessarily completing it ($d = -0.28$). The message from this study is to reduce the use of praise-only feedback during the formative phase if you want the student to focus on the substantive feedback to then improve their writing. In a summative situation, however, there can be praise-only feedback, although more investigation is needed of such praise on subsequent activities in the class (Skipper and Douglas, 2012).

The improvement was greater for university than high school students and this is probably because university instructors were more likely to provide where to next feedback and inviting students to seek additional help. It is not clear why high school teachers are less likely to offer "where to next" feedback, although it is noted they were more likely to request the student seek additional help. Both high school and college students do not seem to mind the source of the feedback, especially the timeliness, accessibility, and quantity of feedback provided by computer-based systems.

The strengths of the study include the large sample size and there was information from a first submission of an essay with formative feedback, then resubmission for summative feedback. The findings invite further study about the role of praise, the possible effects of combinations of forms of feedback (not explored in this study); a major message is the possibilities offered from computer-moderated feedback systems. These systems include both teacher- and automatic-generated feedback, but as important are the facilities and ease for instructors to add inline comments and drag-and-drop comments. The Turnitin Feedback Studio model does not yet provide artificial intelligence provision of "where to next" feedback, but this is well worth investigation and building. The use of a computer-aided system of feedback augmented with teacher-provided feedback does lead to enhanced performance over time.

This study demonstrates that students do appreciate and act upon "where to next" feedback that guides them to enhance their learning and performance, they do not seem to mind whether the feedback is from the teacher via a computer-based feedback tool, and were able, in light of the feedback, to decode and act on the feedback statements.

DATA AVAILABILITY STATEMENT

AUTHOR CONTRIBUTIONS

JH is the first author and conducted the data analysis independent of the co-authors employed by Turnitin, which furnished the dataset. JC, KVG, PW-S, and KW provided information on instructor usage of the Turnitin Feedback Studio product and addressed specific questions of data interpretation that arose during the analysis.

REFERENCES

Arbuckle, J. L. (2011). *IBM SPSS Amos 20 User's Guide*. Wexford, PA: Amos Development Corporation, SPSS Inc.

Azevedo, R., and Bernard, R. M. (1995). A Meta-Analysis of the Effects of Feedback in Computer-Based Instruction. *J. Educ. Comput. Res.* 13 (2), 111–127. doi:10.2190/9lmd-3u28-3a0g-ftqt

Beaumont, C., O'Doherty, M., and Shannon, L. (2011). Reconceptualising Assessment Feedback: A Key to Improving Student Learning?. *Stud. Higher Edu.* 36 (6), 671–687. doi:10.1080/03075071003731135

Brooks, C., Carroll, A., Gillies, R. M., and Hattie, J. (2019). A Matrix of Feedback for Learning. *Aust. J. Teach. Edu.* 44 (4), 13–32. doi:10.14221/ajte.2018v44n4.2

Brown, G. (2009). "The reliability of essay scores: The necessity of rubrics and moderation," in *Tertiary assessment and higher education student outcomes: Policy, practice and research*. Editors L. H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P. M. Johnston and M. Rees, (Wellington, NZ: Ako Aotearoa), 40–48.

Carruthers, C., McCarron, B., Bolan, P., Devine, A., McMahon-Beattie, U., and Burns, A. (2015). 'I like the Sound of that' - an Evaluation of Providing Audio Feedback via the Virtual Learning Environment for Summative Assessment. *Assess. Eval. Higher Edu.* 40 (3), 352–370. doi:10.1080/02602938.2014.917145

Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., et al. (2019). What Makes for Effective Feedback: Staff and Student Perspectives. *Assess. Eval. Higher Edu.* 44 (1), 25–36. doi:10.1080/02602938.2018.1467877

Deci, E. L., Koestner, R., and Ryan, R. M. (1999). A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. *Psychol. Bull.* 125 (6), 627–668. doi:10.1037/0033-2909.125.6.627

Gamlem, S. M., and Smith, K. (2013). Student Perceptions of Classroom Feedback. *Assess. Educ. Principles, Pol. Pract.* 20 (2), 150–169. doi:10.1080/0969594x.2012.749212

Hattie, J., and Timperley, H. (2007). The Power of Feedback. *Rev. Educ. Res.* 77 (1), 81–112. doi:10.3102/003465430298487

Jaehnig, W., and Miller, M. L. (2007). Feedback Types in Programmed Instruction: A Systematic Review. *Psychol. Rec.* 57 (2), 219–232. doi:10.1007/bf03395573

Kluger, A. N., and DeNisi, A. (1996). The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychol. Bull.* 119 (2), 254–284. doi:10.1037/0033-2909.119.2.254

Mandouit, L. (2020). *Investigating How Students Receive, Interpret, and Respond to Teacher Feedback*. Melbourne, Victoria, Australia: Unpublished doctoral dissertation, University of Melbourne.

Pajares, F., and Graham, L. (1998). Formalist Thinking and Language Arts Instruction. *Teach. Teach. Edu.* 14 (8), 855–870. doi:10.1016/s0742-051x(98)80001-2

Sadler, D. R. (1989). Formative Assessment and the Design of Instructional Systems. *Instr. Sci.* 18 (2), 119–144. doi:10.1007/bf00117714

Skipper, Y., and Douglas, K. (2012). Is No Praise Good Praise? Effects of Positive Feedback on Children's and University Students' Responses to Subsequent Failures. *Br. J. Educ. Psychol.* 82 (2), 327–339. doi:10.1111/j.2044-8279.2011.02028.x

Sujee, E., Engelbrecht, A., and Nagel, L. (2015). Effectively Digitizing Communication with Turnitin for Improved Writing in a Multilingual Classroom. *J. Lang. Teach.* 49 (2), 11–31. doi:10.4314/jlt.v49i2.1

Turnitin (2013). *Closing the Gap: What Students Say about Instructor Feedback*. Oakland, CA: Turnitin, LLC. Retrieved from http://go.turnitin.com/what-students-say-about-teacher-feedback?Product=Turnitin&Notification_Language=English&Lead_Origin=Website&source=Website%20-%20Download.

Turnitin (2015). *From Here to There: Students' Perceptions on Feedback, Goals, Barriers, and Effectiveness*. Oakland, CA: Turnitin, LLC. Retrieved from http://go.turnitin.com/paper/student-feedback-goals-barriers.

Turnitin (2014). *Instructor Feedback Writ Large: Student Perceptions on Effective Feedback*. Oakland, CA: Turnitin, LLC. Retrieved from http://go.turnitin.com/paper/student-perceptions-on-effective-feedback.

Van der Kleij, F. M., Feskens, R. C. W., and Eggen, T. J. H. M. (2015). Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes. *Rev. Educ. Res.* 85 (4), 475–511. doi:10.3102/0034654314564881

Watkins, D., Dummer, P., Hawthorne, K., Cousins, J., Emmett, C., and Johnson, M. (2014). Healthcare Students' Perceptions of Electronic Feedback through GradeMark. *JITE:Research* 13, 027–047. doi:10.28945/1945 Retrieved from http://www.jite.org/documents/Vol13/JITEv13ResearchP027-047Watkins0592.pdf.

Wisniewski, B., Zierer, K., and Hattie, J. (2020). The Power of Feedback Revisited: A Meta Analysis of Educational Feedback Research. *Front. Psychol.* 10, 3087. doi:10.3389/fpsyg.2019.03087